

# WO02083953

## Publication Title:

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

## Abstract:

The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

---

Data supplied from the esp@cenet database - <http://ep.espacenet.com>

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number  
WO 02/083953 A1

(51) International Patent Classification<sup>7</sup>: C12Q 1/68, (81) Designated States (national): AE, AG, AL, AM, AT, AU, C07H 21/02, G01N 27/26 AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(21) International Application Number: PCT/US02/11757

(22) International Filing Date: 11 April 2002 (11.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/282,965 11 April 2001 (11.04.2001) US

(71) Applicant (for all designated States except US): PTC THERAPEUTICS, INC. [US/US]; 100 Corporate Court, Middlesex Business Center, South Plainfield, NJ 07080 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): RANDO, Robert [US/US]; 3 Brown Court, Annandale, NJ 08801 (US). WELCH, Ellen [US/US]; 33 Hollow Brook Road, Califon, NJ 07830 (US).

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report  
— with amended claims

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A1

WO 02/083953

(54) Title: METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

(57) Abstract: The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

## METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

5        This application claims the benefit of U.S. Provisional Application No. 60/282,965, filed April 11, 2001, which is incorporated herein by reference in its entirety.

### 1. INTRODUCTION

10      The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of 15 compounds to identify pharmaceutical leads.

### 2. BACKGROUND OF THE INVENTION

20      Protein-nucleic acid interactions are involved in many cellular functions, including transcription, RNA splicing, mRNA decay, and mRNA translation. Readily accessible synthetic molecules that can bind with high affinity to specific sequences of single- or double-stranded nucleic acids have the potential to interfere with these interactions in a controllable way, making them attractive tools for molecular biology and medicine. Successful approaches for blocking function of target nucleic acids include using duplex-forming antisense oligonucleotides (Miller, 1996, *Progress in Nucl. Acid Res. & Mol. Biol.* 52:261-291; Ojwang & Rando, 1999, Achieving antisense inhibition by oligodeoxynucleotides containing N, modified 2'-deoxyguanosine using tumor necrosis factor receptor type 1, METHODS: A Companion to Methods in Enzymology 18:244-251) and peptide nucleic acids ("PNA") (Nielsen, 1999, *Current Opinion in Biotechnology* 10:71-75), which bind to nucleic acids via Watson-Crick base-pairing. Triplex-forming 25 anti-gene oligonucleotides can also be designed (Ping *et al.*, 1997, *RNA* 3:850-860; Aggarwal *et al.*, 1996, *Cancer Res.* 56:5156-5164; U.S. Patent No. 5,650,316), as well as pyrrole-imidazole polyamide oligomers (Gottesfeld *et al.*, 1997, *Nature* 387:202-205; White *et al.*, 1998, *Nature* 391:468-471), which are specific for the major and minor grooves of a 30 double helix, respectively.

In addition to synthetic nucleic acids (*i.e.*, antisense, ribozymes, and triplex-forming molecules), there are examples of natural products that interfere with deoxyribonucleic acid ("DNA") or RNA processes such as transcription or translation. For example, certain carbohydrate-based host cell factors, calicheamicin oligosaccharides, interfere with the sequence-specific binding of transcription factors to DNA and inhibit transcription *in vivo* (Ho *et al.*, 1994, Proc. Natl. Acad. Sci. USA 91:9203-9207; Liu *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93:940-944). Certain classes of known antibiotics have been characterized and were found to interact with RNA. For example, the antibiotic thiostreptone binds tightly to a 60-mer from ribosomal RNA (Cundliffe *et al.*, 1990, in *The Ribosome: Structure, Function & Evolution* (Schlessinger *et al.*, eds.) American Society for Microbiology, Washington, D.C. pp. 479-490). Bacterial resistance to various antibiotics often involves methylation at specific rRNA sites (Cundliffe, 1989, Ann. Rev. Microbiol. 43:207-233). Aminoglycosidic aminocyclitol (aminoglycoside) antibiotics and peptide antibiotics are known to inhibit group I intron splicing by binding to specific regions of the RNA (von Ahsen *et al.*, 1991, Nature (London) 353:368-370). Some of these same aminoglycosides have also been found to inhibit hammerhead ribozyme function (Stage *et al.*, 1995, RNA 1:95-101). In addition, certain aminoglycosides and other protein synthesis inhibitors have been found to interact with specific bases in 16S rRNA (Woodcock *et al.*, 1991, EMBO J. 10:3099-3103). An oligonucleotide analog of the 16S rRNA has also been shown to interact with certain aminoglycosides (Purohit *et al.*, 1994, Nature 370:659-662). A molecular basis for hypersensitivity to aminoglycosides has been found to be located in a single base change in mitochondrial rRNA (Hutchin *et al.*, 1993, Nucleic Acids Res. 21:4174-4179). Aminoglycosides have also been shown to inhibit the interaction between specific structural RNA motifs and the corresponding RNA binding protein. Zapp *et al.* (Cell, 1993, 74:969-978) has demonstrated that the aminoglycosides neomycin B, lividomycin A, and tobramycin can block the binding of Rev, a viral regulatory protein required for viral gene expression, to its viral recognition element in the IIB (or RRE) region of HIV RNA. This blockage appears to be the result of competitive binding of the antibiotics directly to the RRE RNA structural motif.

Single stranded sections of RNA can fold into complex tertiary structures consisting of local motifs such as loops, bulges, pseudoknots, guanosine quartets and turns (Chastain & Tinoco, 1991, Progress in Nucleic Acid Res. & Mol. Biol. 41:131-177; Chow & Bogdan, 1997, Chemical Reviews 97:1489-1514; Rando & Hogan, 1998, *Biologic activity of guanosine quartet forming oligonucleotides in "Applied Antisense Oligonucleotide Technology"* Stein. & Krieg (eds) John Wiley and Sons, New York, pages

335-352). Such structures can be critical to the activity of the nucleic acid and affect functions such as regulation of mRNA transcription, stability, or translation (Weeks & Crothers, 1993; Science 261:1574-1577). The dependence of these functions on the native three-dimensional structural motifs of single-stranded stretches of nucleic acids makes it difficult to identify or design synthetic agents that bind to these motifs using general, simple-to-use sequence-specific recognition rules for the formation of double- and triple-helical nucleic acids used in the design of antisense and ribozyme type molecules.

5 Approaches to screening generally involve competitive assays designed to identify compounds that disrupt the interaction between a target RNA and a physiological, host cell factor(s) that had been previously identified to specifically interact with that particular target RNA. In general, such assays require the identification and characterization of the host cell factor(s) deemed to be required for the function of the target RNA. Both the target RNA and its preselected host cell binding partner are used in a competitive format to identify

10 compounds that disrupt or interfere with the two components in the assay.

15

Citation or identification of any reference in Section 2 of this application is not an admission that such reference is available as prior art to the present invention.

### **3. SUMMARY OF THE INVENTION**

20 The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids including, but not limited to, specific RNA sequences, RNA structural motifs, and/or RNA structural elements. The specific target RNA sequences, RNA structural motifs, and/or RNA structural elements are used as targets for screening small molecules and identifying those that directly bind these specific sequences, motifs, and/or structural elements. For example, methods are described in which

25 a preselected target RNA having a detectable label is used to screen a library of test compounds, preferably under physiologic conditions. Any complexes formed between the target RNA and a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. In particular, the

30 present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the uncomplexed, unlabeled test compounds in the library, and from

35 uncomplexed, labeled target RNA, by a variety of methods, including but not limited to, methods that differentiate changes in the electrophoretic, chromatographic, or thermostable

properties of the complexed target RNA. Such methods include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships (“SAR”) by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. The structure of the test compound attached to the labeled RNA is then determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds, each having an address or identifier, may be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance (“NMR”). The test compounds identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, small organic molecules which interact specifically with target RNA molecules may be useful as lead compounds for the development of therapeutic agents.

The methods described herein for the identification of compounds that directly bind to a particular preselected target RNA are well suited for high-throughput screening. The direct binding method of the invention offers advantages over drug screening systems for competitors that inhibit the formation of naturally-occurring RNA binding protein:target RNA complexes; *i.e.*, competitive assays. The direct binding method of the invention is rapid and can be set up to be readily performed, *e.g.*, by a technician, making it amenable to high throughput screening. The method of the invention also eliminates the bias inherent in the competitive drug screening systems, which require the use of a preselected host cell factor that may not have physiological relevance to the activity of the target RNA. Instead, the methods of the invention are used to identify any compound that can directly bind to specific target RNA sequences, RNA structural motifs, and/or RNA structural elements, preferably under physiologic conditions. As a result, the compounds so identified can inhibit the interaction of the target RNA with any one or more of the native host cell factors (whether known or unknown) required for activity of the RNA *in vivo*.

The present invention may be understood more fully by reference to the detailed description and examples, which are intended to illustrate non-limiting embodiments of the invention.

### 3.1. Definitions

As used herein, a "target nucleic acid" refers to RNA, DNA, or a chemically modified variant thereof. In a preferred embodiment, the target nucleic acid is RNA. A target nucleic acid also refers to tertiary structures of the nucleic acids, such as, but not limited to loops, bulges, pseudoknots, guanosine quartets and turns. A target nucleic acid also refers to RNA elements such as, but not limited to, the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, which are described in Section 5.1. Non-limiting examples of target nucleic acids are presented in Section 5.1 and Section 6.

As used herein, a "library" refers to a plurality of test compounds with which a target nucleic acid molecule is contacted. A library can be a combinatorial library, *e.g.*, a collection of test compounds synthesized using combinatorial chemistry techniques, or a collection of unique chemicals of low molecular weight (less than 1000 daltons) that each occupy a unique three-dimensional space.

As used herein, a "label" or "detectable label" is a composition that is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes (*e.g.*,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^{3}\text{H}$ ), dyes, fluorescent dyes, electron-dense reagents, enzymes and their substrates (*e.g.*, as commonly used in enzyme-linked immunoassays, *e.g.*, alkaline phosphatase and horse radish peroxidase), biotin-streptavidin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. Moreover, a label or detectable moiety can include a "affinity tag" that, when coupled with the target nucleic acid and incubated with a test compound or compound library, allows for the affinity capture of the target nucleic acid along with molecules bound to the target nucleic acid. One skilled in the art will appreciate that a affinity tag bound to the target nucleic acids has, by definition, a complimentary ligand coupled to a solid support that allows for its capture. For example, useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamers, or haptens and proteins for which antisera or monoclonal antibodies are available. The label or detectable moiety is typically bound, either covalently, through a linker or chemical bound, or through ionic, van der Waals or hydrogen bonds to the molecule to be detected.

As used herein, a "dye" refers to a molecule that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means.

As used herein, a "visible dye" refers to a molecule having a chromophore that absorbs radiation in the visible region of the spectrum (*i.e.*, having a wavelength of between about 400 nm and about 700 nm) such that the transmitted radiation is in the visible region and can be detected either visually or by conventional spectroscopic means. As used herein, an "ultraviolet dye" refers to a molecule having a chromophore that absorbs radiation in the ultraviolet region of the spectrum (*i.e.*, having a wavelength of between about 30 nm and about 400 nm). As used herein, an "infrared dye" refers to a molecule having a chromophore that absorbs radiation in the infrared region of the spectrum (*i.e.*, having a wavelength between about 700 nm and about 3,000 nm). A "chromophore" is the network of atoms of the dye that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. One of skill in the art will readily appreciate that although a dye absorbs radiation in one region of the spectrum, it may emit radiation in another region of the spectrum. For example, an ultraviolet dye may emit radiation in the visible region of the spectrum. One of skill in the art will also readily appreciate that a dye can transmit radiation or can emit radiation via fluorescence or phosphorescence.

The phrase "pharmaceutically acceptable salt(s)," as used herein includes but is not limited to salts of acidic or basic groups that may be present in test compounds identified using the methods of the present invention. Test compounds that are basic in nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, *i.e.*, salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (*i.e.*, 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Test compounds that include an amino moiety may form pharmaceutically or cosmetically acceptable salts with various amino acids, in addition to the acids mentioned above. Test compounds that are acidic in nature are capable of forming base salts with various pharmacologically or cosmetically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

5 By "substantially one type of test compound," as used herein, is meant that the assay can be performed in such a fashion that at some point, only one compound need be used in each reaction so that, if the result is indicative of a binding event occurring between the target RNA molecule and the test compound, the test compound can be easily identified.

#### 4. DESCRIPTION OF DRAWINGS

FIG. 1. Gel retardation analysis to detect peptide-RNA interactions. In 20  $\mu$ l reactions containing increasing concentrations of Tat<sub>47-58</sub> peptide (0.1  $\mu$ M, 0.2  $\mu$ M, 0.4  $\mu$ M, 0.8  $\mu$ M, 1.6  $\mu$ M) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90 °C for 2 min and allowed to cool slowly to 24 °C. 10  $\mu$ l of 30% glycerol was added to each sample and applied to a 12% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4 °C in TBE Buffer. 15 Following electrophoresis, the gel was dried and the radioactivity was quantitated with a phosphorimager. The concentration of peptide added is indicated above each lane.

FIG. 2. Gentamicin interacts with an oligonucleotide corresponding to the 16S rRNA. 20  $\mu$ l reactions containing increasing concentrations of gentamicin (1 ng/ml, 10 ng/ml, 100 ng/ml, 1  $\mu$ g/ml, 10  $\mu$ g/ml, 50  $\mu$ g/ml, 500  $\mu$ g/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer, heated at 90 °C for 2 min and allowed to cool slowly to 24 °C. Then 10  $\mu$ l of 30% glycerol was added to each sample and the samples were applied to a 13.5% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4 °C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated using a phosphorimager. The concentration of gentamicin added is indicated above each lane. 25

FIG. 3. The presence of 10 pg/ml gentamicin produces a gel mobility shift in the presence of the 16S rRNA oligonucleotide. 20  $\mu$ l reactions containing increasing concentrations of gentamicin (100 ng/ml, 10 ng/ml, 1 ng/ml, 100 pg/ml, and 10 pg/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer were treated as described for Figure 2. 30

FIG. 4. Gentamicin binding to the 16S rRNA oligonucleotide is weak in the absence of MgCl<sub>2</sub>. Reaction mixtures containing gentamicin (1 mg/ml, 100  $\mu$ g/ml, 35

10  $\mu\text{g/ml}$ , 1  $\mu\text{g/ml}$ , 0.1  $\mu\text{g/ml}$ , and 10  $\text{ng/ml}$ ) were treated as described in Figure 2 except that the TKM buffer does not contain  $\text{MgCl}_2$ .

FIG. 5. Gel retardation analysis to detect peptide-RNA interactions. In reactions 5 containing increasing concentrations of  $\text{Tat}_{47-58}$  peptide (0.1  $\mu\text{M}$ , 0.2  $\mu\text{M}$ , 0.4  $\mu\text{M}$ , 0.8  $\mu\text{M}$ , 1.6  $\mu\text{M}$ ) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90°C for 2 min and allowed to cool slowly to 24°C. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, 10 Pennsylvania). The peaks correspond to the amount of free TAR RNA ("TAR") or the Tat-TAR complex ("Tat-TAR"). The concentration of peptide added is indicated below each lane.

### 5. DETAILED DESCRIPTION OF THE INVENTION

15 The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids, in particular, RNAs, including but not limited to preselected target RNA sequencing structural motifs, or structural elements. Methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds. Any complexes formed between the target RNA and 20 a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. Changes in the physical property of the RNA-test compound complex relative to the target RNA or test compound can be measured by methods such as, but not limited to, methods that detect a change in mobility due to a change in mass, change in charge, or a change in thermostability. Such methods 25 include, but are not limited to; electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in 30 solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the unlabeled, uncomplexed test compounds in the library by a variety of methods capable of differentiating changes in the physical properties of the complexed target RNA. The structure of the test compound 35 attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds,

each having an address or identifier, may be deconvoluted, e.g., by cross-referencing the positive sample to an original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR").

5 Thus, the methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of test compounds, in which the test compounds of the library that specifically bind a preselected target nucleic acid are easily distinguished from non-binding members of the library. The structures of the binding molecules are deciphered from the input library by methods depending on the type of library that is used. 10 The test compounds so identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and lead compounds for development of therapeutics, and the like. Small organic compounds that are identified to 15 interact specifically with the target RNA molecules are particularly attractive candidates as lead compounds for the development of therapeutic agents.

The assay of the invention reduces bias introduced by competitive binding assays which require the identification and use of a host cell factor (presumably essential for modulating RNA function) as a binding partner for the target RNA. The assays of the 20 present invention are designed to detect any compound or agent that binds to the target RNA, preferably under physiologic conditions. Such agents can then be tested for biological activity, without establishing or guessing which host cell factor or factors is required for modulating the function and/or activity of the target RNA.

Section 5.1 describes examples of protein-RNA interactions that are 25 important in a variety of cellular functions and several target RNA elements that can be used to identify test compounds. Compounds that inhibit these interactions by binding to the RNA and successfully competing with the natural protein or host cell factor that endogenously binds to the RNA may be important, e.g., in treating or preventing a disease or abnormal condition, such as an infection or unchecked growth. Section 5.2 describes 30 detectable labels for target nucleic acids that are useful in the methods of the invention. Section 5.3 describes libraries of test compounds. Section 5.4 provides conditions for binding a labeled target RNA to a test compound of a library and detecting RNA binding to a test compound using the methods of the invention. Section 5.5 provides methods for separating complexes of target RNAs bound to a test compound from an unbound RNA. 35 Section 5.6 describes methods for identifying test compounds that are bound to the target RNA. Section 5.7 describes a secondary, biological screen of test compounds identified by

the methods of the invention to test the effect of the test compounds *in vivo*. Section 5.8 describes the use of test compounds identified by the methods of the invention for treating or preventing a disease or abnormal condition in mammals.

5

### **5.1. Biologically Important RNA-Host Cell Factor Interactions**

Nucleic acids, and in particular RNAs, are capable of folding into complex tertiary structures that include bulges, loops, triple helices and pseudoknots, which can provide binding sites for host cell factors, such as proteins and other RNAs. RNA-protein and RNA-RNA interactions are important in a variety cellular functions, including transcription, RNA splicing, RNA stability and translation. Furthermore, the binding of such host cell factors to RNAs may alter the stability and translational efficiency of such RNAs, and accordingly affect subsequent translation. For example, some diseases are associated with protein overproduction or decreased protein function. In this case, the identification of compounds to modulate RNA stability and translational efficiency will be useful to treat and prevent such diseases.

The methods of the present invention are useful for identifying test compounds that bind to target RNA elements in a high throughput screening assay of libraries of test compounds in solution. In particular, the methods of the present invention are useful for identifying a test compound that binds to a target RNA elements and inhibits the interaction of that RNA with one or more host cell factors *in vivo*. The molecules identified using the methods of the invention are useful for inhibiting the formation of a specific bound RNA:host cell factor complexes *in vivo*.

In some embodiments, test compounds identified by the methods of the invention are useful for increasing or decreasing the translation of messenger RNAs (“mRNAs”), *e.g.*, protein production, by binding to one or more regulatory elements in the 5' untranslated region, the 3' untranslated region, or the coding region of the mRNA. Compounds that bind to mRNA can, *inter alia*, increase or decrease the rate of mRNA processing, alter its transport through the cell, prevent or enhance binding of the mRNA to ribosomes, suppressor proteins or enhancer proteins, or alter mRNA stability. Accordingly, compounds that increase or decrease mRNA translation can be used to treat or prevent disease. For example, diseases associated with protein overproduction, such as amyloidosis, or with the production of mutant proteins, such as *Ras*, can be treated or prevented by decreasing translation of the mRNA that codes for the overproduced protein, thus inhibiting production of the protein. Conversely, the symptoms of diseases associated with decreased protein function, such as hemophilia, may be treated by increasing

translation of mRNA coding for the protein whose function is decreased, *e.g.*, factor IX in some forms of hemophilia.

The methods of the invention can be used to identify compounds that bind to mRNAs coding for a variety of proteins with which the progression of diseases in mammals is associated. These mRNAs include, but are not limited to, those coding for amyloid protein and amyloid precursor protein; anti-angiogenic proteins such as angiostatin, endostatin, METH-1 and METH-2; apoptosis inhibitor proteins such as survivin, clotting factors such as Factor IX, Factor VIII, and others in the clotting cascade; collagens; cyclins and cyclin inhibitors, such as cyclin dependent kinases, cyclin D1, cyclin E, WAF1, cdk4 inhibitor, and MTS1; cystic fibrosis transmembrane conductance regulator gene (CFTR); cytokines such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17 and other interleukins; hematopoietic growth factors such as erythropoietin (Epo); colony stimulating factors such as G-CSF, GM-CSF, M-CSF, SCF and thrombopoietin; growth factors such as BDNF, BMP, GGRP, EGF, FGF, GDNF, GGF, HGF, IGF-1, IGF-2, KGF, myotrophin, NGF, OSM, PDGF, somatotrophin, TGF- $\beta$ , TGF- $\alpha$  and VEGF; antiviral cytokines such as interferons, antiviral proteins induced by interferons, TNF- $\alpha$ , and TNF- $\beta$ ; enzymes such as cathepsin K, cytochrome P-450 and other cytochromes, farnesyl transferase, glutathione-S transferases, heparanase, HMG CoA synthetase, N-acetyltransferase, phenylalanine hydroxylase, phosphodiesterase, ras carboxyl-terminal protease, telomerase and TNF converting enzyme; glycoproteins such as cadherins, *e.g.*, N-cadherin and E-cadherin; cell adhesion molecules; selectins; transmembrane glycoproteins such as CD40; heat shock proteins; hormones such as 5- $\alpha$  reductase, atrial natriuretic factor, calcitonin, corticotrophin releasing factor, diuretic hormones, glucagon, gonadotropin, gonadotropin releasing hormone, growth hormone, growth hormone releasing factor, somatotropin, insulin, leptin, luteinizing hormone, luteinizing hormone releasing hormone, parathyroid hormone, thyroid hormone, and thyroid stimulating hormone; proteins involved in immune responses, including antibodies, CTLA4, hemagglutinin, MHC proteins, VLA-4, and kallikrein-kininogen-kinin system; ligands such as CD4; oncogene products such as *sis*, *hst*, protein tyrosine kinase receptors, *ras*, *abl*, *mos*, *myc*, *fos*, *jun*, *H-ras*, *ki-ras*, *c-fms*, *bcl-2*, *L-myc*, *c-myc*, *gip*, *gsp*, and *HER-2*; receptors such as bombesin receptor, estrogen receptor, GABA receptors, growth factor receptors including EGFR, PDGFR, FGFR, and NGFR, GTP-binding regulatory proteins, interleukin receptors, ion channel receptors, leukotriene receptor antagonists, lipoprotein receptors, opioid pain receptors, substance P receptors, retinoic acid and retinoid receptors, steroid receptors, T-cell receptors, thyroid hormone receptors, TNF receptors; tissue

plasminogen activator; transmembrane receptors; transmembrane transporting systems, such as calcium pump, proton pump, Na/Ca exchanger, MRP1, MRP2, P170, LRP, and cMOAT; transferrin; and tumor suppressor gene products such as *APC*, *brca1*, *brca2*, *DCC*, *MCC*, *MTS1*, *NF1*, *NF2*, *nm23*, *p53* and *Rb*. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic. Other target genes include, but are not limited to, those disclosed in Section 5.1 and Section 6.

The methods of the invention can be used to identify mRNA-binding test compounds for increasing or decreasing the production of a protein, thus treating or preventing a disease associated with decreasing or increasing the production of said protein, respectively. The methods of the invention may be useful for identifying test compounds for treating or preventing a disease in mammals, including cats, dogs, swine, horses, goats, sheep, cattle, primates and humans. Such diseases include, but are not limited to, amyloidosis, hemophilia, Alzheimer's disease, atherosclerosis, cancer, gigantism, dwarfism, hypothyroidism, hyperthyroidism, inflammation, cystic fibrosis, autoimmune disorders, diabetes, aging, obesity, neurodegenerative disorders, and Parkinson's disease. Other diseases include, but are not limited to, those described in Section 5.1 and diseases caused by aberrant expression of the genes disclosed in Example 6. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic.

In other embodiments, test compounds identified by the methods of the invention are useful for preventing the interaction of an RNA, such as a transfer RNA ("tRNA"), an enzymatic RNA or a ribosomal RNA ("rRNA"), with a protein or with another RNA, thus preventing, e.g., assembly of an *in vivo* protein-RNA or RNA-RNA complex that is essential for the viability of a cell. The term "enzymatic RNA," as used herein, refers to RNA molecules that are either self-splicing, or that form an enzyme by virtue of their association with one or more proteins, e.g., as in RNase P, telomerase or small nuclear ribonuclear protein particles. For example, inhibition of an interaction between rRNA and one or more ribosomal proteins may inhibit the assembly of ribosomes, rendering a cell incapable of synthesizing proteins. In addition, inhibition of the interaction of precursor rRNA with ribonucleases or ribonucleoprotein complexes (such as RNase P) that process the precursor rRNA prevent maturation of the rRNA and its assembly into ribosomes. Similarly, a tRNA:tRNA synthetase complex may be inhibited by test

compounds identified by the methods of the invention such that tRNA molecules do not become charged with amino acids. Such interactions include, but are not limited to, rRNA interactions with ribosomal proteins, tRNA interactions with tRNA synthetase, RNase P protein interactions with RNase P RNA, and telomerase protein interactions with telomerase RNA.

In other embodiments, test compounds identified by the methods of the invention are useful for treating or preventing a viral, bacterial, protozoan or fungal infection. For example, transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation response region RNA ("TAR RNA"). HIV TAR RNA is a 59-base stem-loop structure located at the 5'-end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a potential binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 upregulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA are useful as anti-HIV therapeutics (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Hamy *et al.*, 1998, *Biochemistry* 37:5086-5095; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

The methods of the invention can be used to identify test compounds to treat or prevent viral, bacterial, protozoan or fungal infections in a patient. In some embodiments, the methods of the invention are useful for identifying compounds that decrease translation of microbial genes by interacting with mRNA, as described above, or for identifying compounds that inhibit the interactions of microbial RNAs with proteins or other ligands that are essential for viability of the virus or microbe. Examples of microbial target RNAs useful in the present invention for identifying antiviral, antibacterial, anti-protozoan and anti-fungal compounds include, but are not limited to, general antiviral and anti-inflammatory targets such as mRNAs of INF $\alpha$ , INF $\gamma$ , RNase L, RNase L inhibitor protein, PKR, tumor necrosis factor, interleukins 1-15, and IMP dehydrogenase; internal ribosome entry sites; HIV-1 CT rich domain and RNase H mRNA; HCV internal ribosome entry site (required to direct translation of HCV mRNA), and the 3'-untranslated tail of HCV genomes; rotavirus NSP3 binding site, which binds the protein NSP3 that is required for rotavirus mRNA translation; HBV epsilon domain; Dengue virus 5' and 3' untranslated regions, including IRES; INF $\alpha$ , INF $\beta$  and INF $\gamma$ ; plasmodium falciparum mRNAs; the 16S

ribosomal subunit ribosomal RNA and the RNA component of RNase P of bacteria; and the RNA component of telomerase in fungi and cancer cells. Other target viral and bacterial mRNAs include, but are not limited to, those disclosed in Section 6.

5 One of skill in the art will appreciate that, although such target RNAs are functionally conserved in various species (e.g., from yeast to humans), they exhibit nucleotide sequence and structural diversity. Therefore, inhibition of, for example, yeast telomerase by an anti-fungal compound identified by the methods of the invention might not interfere with human telomerase and normal human cell proliferation.

10 Thus, the methods of the invention can be used to identify test compounds that interfere with one or more target RNA interactions with host cell factors that are important for cell growth or viability, or essential in the life cycle of a virus, a bacterium, a protozoa or a fungus. Such test compounds and/or congeners that demonstrate desirable biologic and pharmacologic activity can be administered to a patient in need thereof in order 15 to treat or prevent a disease caused by viral, bacterial, protozoan, or fungal infections. Such diseases include, but are not limited to, HIV infection, AIDS, human T-cell leukemia, SIV infection, FIV infection, feline leukemia, hepatitis A, hepatitis B, hepatitis C, Dengue fever, malaria, rotavirus infection, severe acute gastroenteritis, diarrhea, encephalitis, hemorrhagic fever, syphilis, legionella, whooping cough, gonorrhea, sepsis, influenza, pneumonia, tinea 20 infection, candida infection, and meningitis.

25 Non-limiting examples of RNA elements involved in the regulation of gene expression, *i.e.*, mRNA stability, translational efficiency via translational initiation and ribosome assembly, *etc.*, include the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, as discussed below.

### 5.1.1. HIV TAR Element

Transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation 30 response region RNA ("TAR RNA"), a 59-base stem-loop structure located at the 5' end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a useful binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat- 35 TAR RNA complex involved in HIV-1 up-regulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA

can be useful as anti-HIV therapeutics (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Hamy *et al.*, 1998, Biochemistry 37:5086-5095; Mei *et al.*, 1998, Biochemistry 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

5

### 5.1.2. Internal Ribosome Entry Site ("IRES")

Internal ribosome entry sites ("IRES") are found in the 5' untranslated regions ("5' UTR") of several mRNAs, and are thought to be involved in the regulation of translational efficiency. When the IRES element is present on an mRNA downstream of a 10 translational stop codon, it directs ribosomal re-entry (Ghattas *et al.*, 1991, Mol. Cell. Biol. 11:5848-5959), which permits initiation of translation at the start of a second open reading frame.

As reviewed by Jang *et al.*, a large segment of the 5' nontranslated region, 15 approximately 400 nucleotides in length, promotes internal entry of ribosomes independent of the non-capped 5' end of picornavirus mRNAs (mammalian plus-strand RNA viruses whose genomes serve as mRNA). This 400 nucleotide segment (IRES), maps approximately 200 nt down-stream from the 5' end and is highly structured. IRES elements of different picornaviruses, although functionally similar *in vitro* and *in vivo*, are not identical in sequence or structure. However, IRES elements of the genera entero- and 20 rhinoviruses, on the one hand, and cardio- and aphthoviruses, on the other hand, reveal similarities corresponding to phylogenetic kinship. All IRES elements contain a conserved Yn-Xm-AUG unit (Y, pyrimidine; X, nucleotide) which appears essential for IRES function. The IRES elements of cardio-, entero- and aphthoviruses bind a cellular protein, p57. In the case of cardioviruses, the interaction between a specific stem-loop of the IREs is 25 essential for translation *in vitro*. The IRES elements of entero- and cardioviruses also bind the cellular protein, p52, but the significance of this interaction remains to be shown. The function of p57 or p52 in cellular metabolism is unknown. Since picornaviral IRES elements function *in vivo* in the absence of any viral gene products, it is speculated that 30 IRES-like elements may also occur in specific cellular mRNAs releasing them from cap-dependent translation (Jang *et al.*, 1990, Enzyme 44(1-4):292-309).

### 5.1.3. "Slippery Site"

Programmed, or directed, ribosomal frameshifting, when ribosomes shift 35 from one translation reading frame to another and synthesize two viral proteins from a single viral mRNA, is directed by a unique site in viral mRNAs called the "slippery site." The slippery site directs ribosomal frameshifting in the -1 or +1 direction that causes the

ribosome to slip by one base in the 5' direction thereby placing the ribosome in the new reading frame to produce a new protein.

Programmed, or directed, ribosomal frameshifting is of particular value to viruses that package their plus strands, as it eliminates the need to splice their mRNAs and reduces the risk of packaging defective genomes and regulates the ratio of viral proteins synthesized. Examples of programmed translational frameshifting (both +1 and -1 shifts) have been identified in ScV systems (Lopinski *et al.*, 2000, Mol. Cell. Biol. 20(4):1095-103, retroviruses (Falk *et al.*, 1993, J. Virol. 67:273-6277; Jacks & Varmus, 1985, Science 230:1237-1242; Morikawa & Bishop, 1992, Virology 186:389-397; Nam *et al.*, 1993, J. Virol. 67:196-203); coronaviruses (Brierley *et al.*, 1987, EMBO J. 6:3779-3785; Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842); giardiaviruses, which are also members of the Totiviridae (Wang *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:8595-8599); two bacterial genes (Blinkowa & Walker, 1990, Nucleic Acids Res., 18:1725-1729; Craigen & Caskey, 1986, Nature 322:273); bacteriophage genes (Condron *et al.*, 1991, Nucleic Acids Res. 19:5607-5612); astroviruses (Marczinke *et al.*, 1994, J. Virol. 68:5588-5595); the yeast EST3 gene (Lundblad & Morris, 1997, Curr. Biol. 7:969-976); and the rat, mouse, Xenopus, and Drosophila ornithine decarboxylase antizymes (Matsufuji *et al.*, 1995, Cell 80:51-60); and a significant number of cellular genes (Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842).

Drugs targeted to ribosomal frameshifting minimize the problem of virus drug resistance because this strategy targets a host cellular process rather than one introduced into the cell by the virus, which minimizes the ability of viruses to evolve drug-resistant mutants. Compounds that target the RNA elements involved in regulating programmed frameshifting should have several advantages, including (a) any selective pressure on the host cellular translational machinery to adapt to the drugs would have to occur at the host evolutionary time scale, which is on the order of millions of years, (b) ribosomal frameshifting is not used to express any host proteins, and (c) altering viral frameshifting efficiencies by modulating the activity of a host protein minimizing the likelihood that the virus will acquire resistance to such inhibition by mutations in its own genome.

#### 5.1.4. Instability Elements

“Instability elements” may be defined as specific sequence elements that promote the recognition of unstable mRNAs by cellular turnover machinery. Instability

elements have been found within mRNA protein coding regions as well as untranslated regions.

5 Altering the control of stability of normal mRNAs may lead to disease. The alteration of mRNA stability has been implicated in diseases such as, but not limited to, cancer, immune disorders, heart disease, and fibrotic disorders.

There are several examples of mutations that delete instability elements which then result in stabilization of mRNAs that may be involved in the onset of cancer. In Burkitt's lymphoma, a portion of the *c-myc* proto-oncogene is translocated to an Ig locus, 10 producing a form of the *c-myc* mRNA that is five times more stable (see, e.g., Kapstein *et al.*, 1996, *J. Biol. Chem.* 271(31):18875-84). The highly oncogenic *v-fos* mRNA lacks the 3' UTR adenylate uridylate rich element ("ARE") that is found in the more labile and weakly oncogenic *c-fos* mRNA (see, e.g., Schiavi *et al.*, 1992, *Biochim Biophys Acta*, 1114(2-3):95-106). Differences between the benign cervical lesions brought about by 15 nonintegrated circular human papillomavirus type 16 and its integrated form, that lacks the 3' UTR ARE and correlates with cervical carcinomas, may be a consequence of stabilizing the E6/E7 transcripts encoding oncogenic proteins. Integration of the virus results in deletion of the ARE instability element, resulting in stabilization of the transcripts and over-expression of the proteins (see, e.g., Jeon & Lambert, 1995, *Proc. Natl. Acad. Sci. USA* 92(5):1654-8). Deletion of AREs from the 3' UTR of the IL-2 and IL-3 genes promotes 20 increased stabilization of these mRNAs, high expression of these proteins, and leads to the formation of cancerous cells (see, e.g., Stoecklin *et al.*, 2000, *Mol. Cell. Biol.* 20(11):3753-63).

Mutations in trans-acting factors involved in mRNA turnover may also 25 promote cancer. In monocytic tumors, the lymphokine GM-CSF mRNA is specifically stabilized as a consequence of an oncogenic lesion in a trans-acting factor that controls mRNA turnover rates. Furthermore, the normally unstable IL-3 transcript is inappropriately long-lived in mast tumor cells. Similarly, the labile GM-CSF mRNA is greatly stabilized in bladder carcinoma cells. See, e.g., Bickel *et al.*, 1990, *J. Immunol.* 145(3):840-5.

30 The immune system is regulated by a large number of regulatory molecules that either activate or inhibit the immune response. It has now been clearly demonstrated that stability of the transcripts encoding these proteins are highly regulated. Altered regulation of these molecules leads to mis-regulation of this process and can result in drastic medical consequences. For example, recent results using transgenic mice have shown that 35 mis-regulation of the stability of the important modulator TNF $\alpha$  mRNA leads to diseases

such as, but not limited to, rheumatoid arthritis and a Crohn's-like liver disease. *See, e.g.*, Clark, 2000, *Arthritis Res.* 2(3):172-4.

Smooth muscle in the heart is modulated by the  $\beta$ -adrenergic receptor, which 5 in turn responds to the sympathetic neurotransmitter norepinephrine and the adrenal hormone epinephrine. Chronic heart failure is characterized by impairment of smooth muscle cells, which results, in part, from the more rapid decay of the  $\beta$ -adrenergic receptor mRNA. *See, e.g.*, Ellis & Frielle, 1999, *Biochem. Biophys. Res. Commun.* 258(3):552-8.

A large number of diseases result from over-expression of collagen. For 10 example, cirrhosis results from damage to the liver as a consequence of cancer, viral infection, or alcohol abuse. Such damage causes mis-regulation of collagen expression, leading to the formation of large collagen deposits. Recent results indicate that the sizeable increase in collagen expression is largely attributable to stabilization of its mRNA. *See, e.g.*, Lindquist *et al.*, 2000, *Am. J. Physiol. Gastrointest. Liver Physiol.* 279(3):G471-6.

15

### 5.1.5. Adenylate Uridylate-rich Elements (“ARE”)

Adenylate uridylate-rich elements (“ARE”) are found in the 3' untranslated regions (“3' UTR”) of several mRNAs, and involved in the turnover of mRNAs, such as but not limited to transcription factors, cytokines, and lymphokines. AREs may function both 20 as stabilizing and destabilizing elements. ARE mRNAs are classified into five groups, depending on sequence (Bakheet *et al.*, 2001, *Nucl. Acids Res.* 29(1):246-254). An ongoing database at the web site <http://rc.kfshrc.edu.sa/ared> contains ARE-containing mRNAs and their cluster groups, which is incorporated by reference in its entirety. The ARE motifs are classified as follows:

25	Group I Cluster	(AUUUAUUUAUUUAUUUAUUA)	SEQ ID NO: 1
	Group II Cluster	(AUUUAUUUAUUUAUUU) stretch	SEQ ID NO: 2
	Group III Cluster	(WAUUUAUUUAUUUAW) stretch	SEQ ID NO: 3
	Group IV Cluster	(WWAUUUAUUUAWW) stretch	SEQ ID NO: 4
30	Group V Cluster	(WWWWAUUUAWWWW) stretch	SEQ ID NO: 5

The ARE-mRNAs were clustered into five groups containing five, four, three and two pentameric repeats, while the last group contains only one pentamer within the 13-bp ARE pattern. Functional categories were assigned whenever possible according to NCBI-COG functional annotation (Tatusov *et al.*, 2001, *Nucleic Acids Research*, 29(1): 35 22-28), in addition to the categories: inflammation, immune response, development/differentiation, using an extensive literature search.

Group I contains many secreted proteins including GM-CSF, IL-1, IL-11, IL-12 and Gro- $\beta$  that affect the growth of hematopoietic and immune cells (Witsell & Schook, 1992, Proc. Natl Acad. Sci. USA, 89:4754–4758). Although TNF $\alpha$  is both a pro-inflammatory and anti-tumor protein, there is experimental evidence that it can act as a growth factor in certain leukemias and lymphomas (Liu *et al.*, 2000, J. Biol. Chem. 275:21086–21093).

Unlike Group I, Groups II–V contain functionally diverse gene families comprising immune response, cell cycle and proliferation, inflammation and coagulation, angiogenesis, metabolism, energy, DNA binding and transcription, nutrient transportation and ionic homeostasis, protein synthesis, cellular biogenesis, signal transduction, and apoptosis (Bakheet *et al.*, 2001, Nucl. Acids Res. 29(1):246-254).

Several groups have described ARE-binding proteins that influence the ARE-mRNA stability. Among the well-characterized proteins are the mammalian homologs of ELAV (embryonic lethal abnormal vision) proteins including AUF1, HuR and He1-N2 (Zhang *et al.*, 1993, Mol. Cell. Biol. 13:7652–7665; Levine *et al.*, 1993, Mol. Cell. Biol. 13:3494–3504; Ma *et al.*, 1996, J. Biol. Chem. 271:8144–8151). The zinc-finger protein tristetraprolin has been identified as another ARE-binding protein with destabilizing activity on TNF $\alpha$ , IL-3 and GM-CSF mRNAs (Stoecklin *et al.*, 2000, Mol. Cell. Biol. 20:3753–3763; Carballo *et al.*, 2000, Blood 95:1891–1899).

Since ARE-containing genes are clearly important in biological systems, including but not limited to a number of the early response genes that regulate cell proliferation and responses to exogenous agents, the identification of compounds that bind to one or more of the ARE clusters and potentially modulate the stability of the target RNA can potentially be of value as a therapeutic.

### 5.2. Detectably Labeled Target RNAs

Target nucleic acids, including but not limited to RNA and DNA, useful in the methods of the present invention have a label that is detectable via conventional spectroscopic means or radiographic means. Preferably, target nucleic acids are labeled with a covalently attached dye molecule. Useful dye-molecule labels include, but are not limited to, fluorescent dyes, phosphorescent dyes, ultraviolet dyes, infrared dyes, and visible dyes. Preferably, the dye is a visible dye.

Useful labels in the present invention can include, but are not limited to, spectroscopic labels such as fluorescent dyes (*e.g.*, fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green<sup>TM</sup>, rhodamine and derivatives (*e.g.*,

Texas red, tetramethylrhodamine isothiocyanate (TRITC), bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, *etc.*), digoxigenin, biotin, phycoerythrin, AMCA, CyDye™, and the like), radiolabels (*e.g.*,  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ ,  $^{32}\text{P}$ ,  $^{33}\text{P}$ , *etc.*), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase *etc.*), spectroscopic colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.* polystyrene, polypropylene, latex, *etc.*) beads, or nanoparticles – nanoclusters of inorganic ions with defined dimension from 0.1 to 1000 nm. Useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamer-streptavidin, or haptens and proteins for which antisera or monoclonal antibodies are available. The label may be coupled directly or indirectly to a component of the detection assay (*e.g.*, the detection reagent) according to methods well known in the art. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

In one embodiment, nucleic acids that are labeled at one or more specific locations are chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (*see, e.g.*, Caruthers *et al.*, U.S. Pat. Nos. 4,458,066 and 4,415,732; Caruthers *et al.*, 1982, *Genetic Engineering* 4:1-17; *Users Manual Model 392 and 394 Polynucleotide Synthesizers*, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, *et al.*, 1997, *Biochemistry*, 36:6033-6045). The phosphoramidite method of polynucleotide synthesis is the preferred method because of its efficient and rapid coupling and the stability of the starting materials. The synthesis is performed with the growing polynucleotide chain attached to a solid support, such that excess reagents, which are generally in the liquid phase, can be easily removed by washing, decanting, and/or filtration, thereby eliminating the need for purification steps between synthesis cycles.

The following briefly describes illustrative steps of a typical polynucleotide synthesis cycle using the phosphoramidite method. First, a solid support to which is attached a protected nucleoside monomer at its 3' terminus is treated with acid, *e.g.*, trichloroacetic acid, to remove the 5'-hydroxyl protecting group, freeing the hydroxyl group for a subsequent coupling reaction. After the coupling reaction is completed an activated intermediate is formed by contacting the support-bound nucleoside with a protected nucleoside phosphoramidite monomer and a weak acid, *e.g.*, tetrazole. The weak acid protonates the nitrogen atom of the phosphoramidite forming a reactive intermediate.

Nucleoside addition is generally complete within 30 seconds. Next, a capping step is performed, which terminates any polynucleotide chains that did not undergo nucleoside addition. Capping is preferably performed using acetic anhydride and 1-methylimidazole.

5 The phosphite group of the internucleotide linkage is then converted to the more stable phosphotriester by oxidation using iodine as the preferred oxidizing agent and water as the oxygen donor. After oxidation, the hydroxyl protecting group of the newly added nucleoside is removed with a protic acid, *e.g.*, trichloroacetic acid or dichloroacetic acid, and the cycle is repeated one or more times until chain elongation is complete. After

10 synthesis, the polynucleotide chain is cleaved from the support using a base, *e.g.*, ammonium hydroxide or t-butyl amine. The cleavage reaction also removes any phosphate protecting groups, *e.g.*, cyanoethyl. Finally, the protecting groups on the exocyclic amines of the bases and any protecting groups on the dyes are removed by treating the polynucleotide solution in base at an elevated temperature, *e.g.*, at about 55°C. Preferably

15 the various protecting groups are removed using ammonium hydroxide or t-butyl amine.

Any of the nucleoside phosphoramidite monomers can be labeled using standard phosphoramidite chemistry methods (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002; Ojwang *et al.*, 1997, Biochemistry. 36:6033-6045 and references cited therein). Dye molecules useful for covalently coupling to phosphoramidites preferably comprise a primary hydroxyl group that is not part of the dye's chromophore. Illustrative dye molecules include, but are not limited to, disperse dye CAS 4439-31-0, disperse dye CAS 6054-58-6, disperse dye CAS 4392-69-2 (Sigma-Aldrich, St. Louis, MO), disperse red, and 1-pyrenebutanol (Molecular Probes, Eugene, OR). Other dyes useful for coupling to phosphoramidites will be apparent to those of skill in the art, such as fluorescein, cy3, and cy5 fluorescent dyes, and may be purchased from, *e.g.*, Sigma-Aldrich, St. Louis, MO or Molecular Probes, Inc., Eugene, OR.

In another embodiment, dye-labeled target RNA molecules are synthesized enzymatically using *in vitro* transcription (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). In this embodiment, a template DNA is denatured by heating to about 90°C and an oligonucleotide primer is annealed to the template DNA, for example by slow-cooling the mixture of the denatured template and the primer from about 90°C to room temperature. A mixture of ribonucleoside-5'-triphosphates capable of supporting template-directed enzymatic extension of the primed template (*e.g.*, a mixture including GTP, ATP, CTP, and UTP), including one or more dye-labeled

30 ribonucleotides (Sigma-Aldrich, St. Louis, MO), is added to the primed template. Next, a polymerase enzyme is added to the mixture under conditions where the polymerase enzyme

35

is active, which are well-known to those skilled in the art. A labeled polynucleotide is formed by the incorporation of the labeled ribonucleotides during polymerase-mediated strand synthesis.

5 In yet another embodiment of the invention, nucleic acid molecules are end-labeled after their synthesis. Methods for labeling the 5'-end of an oligonucleotide include but are by no means limited to: (i) periodate oxidation of a 5'-to-5'-coupled ribonucleotide, followed by reaction with an amine-reactive label (Heller & Morisson, 1985, in *Rapid Detection and Identification of Infectious Agents*, D.T. Kingsbury and S. Falkow, eds., pp. 10 245-256, Academic Press); (ii) condensation of ethylenediamine with 5'-phosphorylated polynucleotide, followed by reaction with an amine reactive label (Morrison, European Patent Application 232 967); (iii) introduction of an aliphatic amine substituent using an aminohexyl phosphite reagent in solid-phase DNA synthesis, followed by reaction with an amine reactive label (Cardullo *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:8790-8794); and 15 (iv) introduction of a thiophosphate group on the 5'-end of the nucleic acid, using phosphatase treatment followed by end-labeling with ATP-<sup>32</sup>S and kinase, which reacts specifically and efficiently with maleimide-labeled fluorescent dyes (Czworkowski *et al.*, 1991, Biochem. 30:4821-4830).

20 A detectable label should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a 25 covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site. Accordingly, if the region of the target nucleic acid that binds to a host cell factor is known, a detectable label is preferably incorporated into the nucleic acid molecule at one or more positions that are spatially or sequentially remote from the binding region.

25 After synthesis, the labeled target nucleic acid can be purified using standard techniques known to those skilled in the art (see Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the target nucleic acid and the method of its synthesis, such purification techniques include, but 30 are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification. After purification, the target RNA is refolded into its native conformation, preferably by heating to approximately 85-95°C and slowly cooling to room temperature in a buffer, *e.g.*, a buffer comprising about 50 mM Tris-HCl, pH 8 and 100 mM NaCl.

35 In another embodiment, the target nucleic acid can also be radiolabeled. A radiolabel, such as, but not limited to, an isotope of phosphorus, sulfur, or hydrogen, may be

incorporated into a nucleotide, which is added either after or during the synthesis of the target nucleic acid. Methods for the synthesis and purification of radiolabeled nucleic acids are well known to one of skill in the art. See, e.g., Sambrook *et al.*, 1989, in *Molecular Cloning: A Laboratory Manual*, pp 10.2-10.70, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties.

5 In another embodiment, the target nucleic acid can be attached to an inorganic nanoparticle. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to 10  $\text{Ag}_2\text{S}$ ,  $\text{ZnS}$ ,  $\text{CdS}$ ,  $\text{CdTe}$ ,  $\text{Au}$ , or  $\text{TiO}_2$ . Nanoparticles have unique optical, electronic and catalytic properties relative to bulk materials which can be adjusted according to the size of the particle. Methods for the attachment of nucleic acids are well known to one of skill in the art (see, e.g., Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40: 4129-4158, International Patent Publication WO/0218643, and the references cited therein, the disclosures of which are 15 hereby incorporated by reference in their entireties).

### **5.3. Libraries of Small Molecules**

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds. In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a 20 phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as  $\alpha$ -amino phosphoric acids and  $\alpha$ -amino phosphoric acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, 25 synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used.

30 In a preferred embodiment, the combinatorial libraries are small organic molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and diazepindiones. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; 35 vinyllogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries.

Combinatorial libraries are themselves commercially available (see, e.g., Advanced ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow Russia; BioFocus plc, Sittingbourne, UK; Bionet Research (A division of Key Organics Limited ), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California.; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SIDDCO Inc, Tucson, Arizona; TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle, 2001, *J. Comb. Chem.* 3:477-517; Hall *et al.*, 2001, *J. Comb. Chem.* 3:125-150; Dolle, 2000, *J. Comb. Chem.* 2:383-433; Dolle, 1999, *J. Comb. Chem.* 1:235-282). The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry. Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A.W. Czarnik and S.H. Dewitt, eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 1998; *Organic Synthesis on Solid Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses*, Vol. 1, A.W. Czarnik, ed., Wiley Interscience, 2001.

Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No. 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No. 6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby incorporated by reference in their entirety. These patents describe synthesis apparatuses capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid phase synthesis of combinatorial compound libraries, liquid phase synthesis does not require the use of specialized protocols for monitoring the individual steps of a multistep solid phase synthesis (Egner *et al.*, 1995, *J. Org. Chem.* 60:2652; Anderson *et al.*, 1995, *J. Org. Chem.* 60:2650; Fitch *et al.*, 1994, *J. Org. Chem.* 59:7955; Look *et al.*, 1994, *J. Org. Chem.* 49:7588; Metzger *et al.*, 1993, *Angew. Chem., Int. Ed. Engl.* 32:894; Youngquist *et al.*, 1994, *Rapid Commun. Mass Spect.* 8:77; Chu *et al.*, 1995, *J. Am. Chem. Soc.* 117:5419; Brummel *et al.*, 1994, *Science* 264:399; Stevanovic *et al.*, 1993, *Bioorg. Med. Chem. Lett.* 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing solid supports during the synthesis, is used to synthesize a library of compounds on solid supports (see Lam *et al.*, 1997, *Chem. Rev.* 97:41-448; Ohlmeyer *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (see, e.g., Nefzi *et al.*, 1997, *Chem. Rev.* 97:449-472 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their entirety).

As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, cotton, plastic beads, polystyrene beads, alumina gels, and polysaccharides. A suitable solid support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p-methylbenzhydrylamine (pMBHA) resin (Peptides International, Louisville, KY), polystyrenes (e.g., PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), including chloromethylpolystyrene, hydroxymethylpolystyrene and

aminomethylpolystyrene, poly (dimethylacrylamide)-grafted styrene co-divinyl-benzene (e.g., POLYHIPE resin, obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (e.g., 5 TENTAGEL or ARGOGEL, Bayer, Tubingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Biosearch, California), or Sepharose (Pharmacia, Sweden).

In one embodiment, the solid phase support is suitable for *in vivo* use, *i.e.*, it can serve as a carrier or support for administration of the test compound to a patient (e.g., 10 TENTAGEL, Bayer, Tubingen, Germany). In a particular embodiment, the solid support is palatable and/or orally ingestable.

In some embodiments of the present invention, compounds can be attached 15 to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For 20 example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions.

In another embodiment, the combinatorial compound libraries can be 25 assembled *in situ* using dynamic combinatorial chemistry as described in European Patent Application 1,118,359 A1 to Lehn; Huc & Nguyen, 2001, Comb. Chem. High Throughput. Screen. 4:53-74; Lehn and Eliseev, 2001, Science 291:2331-2332; Cousins *et al.* 2000, Curr. Opin. Chem. Biol. 4: 270-279; and Karan & Miller, 2000, Drug. Disc. Today 5:67-75 which are incorporated by reference in their entirety.

Dynamic combinatorial chemistry uses non-covalent interaction with a 30 target biomolecule, including but not limited to a protein, RNA, or DNA, to favor assembly of the most tightly binding molecule that is a combination of constituent subunits present as a mixture in the presence of the biomolecule. According to the laws of thermodynamics, when a collection of molecules is able to combine and recombine at equilibrium through reversible chemical reactions in solution, molecules, preferably one molecule, that bind 35 most tightly to a templating biomolecule will be present in greater amount than all other possible combinations. The reversible chemical reactions include, but are not limited to, imine, acyl-hydrazone, amide, acetal, or ester formation between carbonyl-containing compounds and amines, hydrazines, or alcohols; thiol exchange between disulfides; alcohol

exchange in borate esters; Diels-Alder reactions; thermal- or photoinduced sigmatropic or electrocyclic rearrangements; or Michael reactions.

In the preferred embodiment of this technique, the constituent components of the dynamic combinatorial compound library are allowed to combine and reach equilibrium in the absence of the target RNA and then incubated in the presence of the target RNA, preferably at physiological conditions, until a second equilibrium is reached. The second, perturbed, equilibrium (the so-called "templated mixture") can, but need not necessarily, be fixed by a further chemical transformation, including but not limited to reduction, oxidation, hydrolysis, acidification, or basification, to prevent restoration of the original equilibrium when the dynamical combinatorial compound library is separated from the target RNA.

In the preferred embodiment of this technique, the predominant product or products of the templated dynamic combinatorial library can be separated from the minor products and directly identified. In another embodiment, the identity of the predominant product or products can be identified by a deconvolution strategy involving preparation of derivative dynamic combinatorial libraries, as described in European Patent Application 1,118,359 A1, which is incorporated by reference in their entirety, whereby each component of the mixture is, preferably one-by-one but possibly group-wise, left out of the mixture and the ability of the derivative library mixture at chemical equilibrium to bind the target RNA is measured. The components whose removal most greatly reduces the ability of the derivative dynamic combinatorial library to bind the target RNA are likely the components of the predominant product or products in the original dynamic combinatorial library.

25

#### **5.4. Library Screening**

After a target nucleic acid, such as but not limited to RNA or DNA, is labeled and a test compound library is synthesized or purchased or both, the labeled target nucleic acid is used to screen the library to identify test compounds that bind to the nucleic acid. Screening comprises contacting a labeled target nucleic acid with an individual, or small group, of the components of the compound library. Preferably, the contacting occurs in an aqueous solution, and most preferably, under physiologic conditions. The aqueous solution preferably stabilizes the labeled target nucleic acid and prevents denaturation or degradation of the nucleic acid without interfering with binding of the test compounds. The aqueous solution can be similar to the solution in which a complex between the target RNA and its corresponding host cell factor (if known) is formed *in vitro*. For example, TK

buffer, which is commonly used to form Tat protein-TAR RNA complexes *in vitro*, can be used in the methods of the invention as an aqueous solution to screen a library of test compounds for TAR RNA binding compounds.

5 The methods of the present invention for screening a library of test compounds preferably comprise contacting a test compound with a target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions. The aqueous solution optionally further comprises non-specific nucleic acids, such as, but not 10 limited to, DNA; yeast tRNA; salmon sperm DNA; homoribopolymers such as, but not limited to, poly IC, polyA, polyU, and polyC; and non-specific RNA. The non-specific RNA may be an unlabeled target nucleic acid having a mutation at the binding site, which renders the unlabeled nucleic acid incapable of interacting with a test compound at that site. For example, if dye-labeled TAR RNA is used to screen a library, unlabeled TAR RNA 15 having a mutation in the uracil 23/cytosine 24 bulge region may also be present in the aqueous solution. Without being bound by any theory, the addition of unlabeled RNA that is essentially identical to the dye-labeled target RNA except for a mutation at the binding site might minimize interactions of other regions of the dye-labeled target RNA with test compounds or with the solid support and prevent false positive results.

20 The solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. The pH of the solution typically ranges from about 5 to about 8, preferably from about 6 to about 8, most preferably from about 6.5 to about 8. A variety of buffers may be used to achieve the desired pH. Suitable buffers include, but are not limited to, Tris, Mes, Bis-Tris, Ada, Aces, Pipes, Mopso, Bis-Tris propane, Bes, 25 Mops, Tes, Hepes, Dipso, Mobs, Tapso, Trizma, Heppso, Popso, TEA, Epps, Tricine, Gly-Gly, Bicine, and sodium-potassium phosphate. The buffering agent comprises from about 10 mM to about 100 mM, preferably from about 25 mM to about 75 mM, most preferably from about 40 mM to about 60 mM buffering agent. The pH of the aqueous solution can be optimized for different screening reactions, depending on the target RNA used and the 30 types of test compounds in the library, and therefore, the type and amount of the buffer used in the solution can vary from screen to screen. In a preferred embodiment, the aqueous solution has a pH of about 7.4, which can be achieved using about 50 mM Tris buffer.

35 In addition to an appropriate buffer, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl<sub>2</sub>. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl<sub>2</sub>. Without

5 being bound by any theory, Applicant has found that a combination of KCl, NaCl, and MgCl<sub>2</sub> stabilizes the target RNA such that most of the RNA is not denatured or digested over the course of the screening reaction. The optional concentration of each salt used in the aqueous solution is dependent on the particular target RNA used and can be determined using routine experimentation.

10 The solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant. Without being bound by any theory, a small amount of detergent or surfactant in the solution might reduce non-specific binding of the target RNA to the solid support and control aggregation and increase stability of target RNA molecules.

15 Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alkyl amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl  $\alpha$ -D-glucopyranoside, n-decyl  $\beta$ -D-maltopyranoside, n-dodecyl  $\beta$ -D-maltoside, n-octyl  $\beta$ -D-glucopyranoside, sorbitan esters, n-tetradecyl  $\beta$ -D-maltoside, octylphenoxy polyethoxyethanol (Nonidet P-40), nonylphenoxy polyethoxyethanol (NP-40), and tritons.

20 Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol. More preferably, the detergent, if present, is Triton X-100 and present in an amount of about 0.1% (w/v).

25 Non-specific binding of a labeled target nucleic acid to test compounds can be further minimized by treating the binding reaction with one or more blocking agents. In one embodiment, the binding reactions are treated with a blocking agent, e.g., bovine serum albumin ("BSA"), before contacting with the labeled target nucleic acid. In another embodiment, the binding reactions are treated sequentially with at least two different blocking agents. This blocking step is preferably performed at room temperature for from about 0.5 to about 3 hours. In a subsequent step, the reaction mixture is further treated with unlabeled RNA having a mutation at the binding site. This blocking step is preferably performed at about 4°C for from about 12 hours to about 36 hours before addition of the dye-labeled target RNA. Preferably, the solution used in the one or more blocking steps is

substantially similar to the aqueous solution used to screen the library with the dye-labeled target RNA, *e.g.*, in pH and salt concentration.

Once contacted, the mixture of labeled target nucleic acid and the test compound is preferably maintained at 4°C for from about 1 day to about 5 days, preferably from about 2 days to about 3 days with constant agitation. To identify the reactions in which binding to the labeled target nucleic acid occurred, after the incubation period, bound from free compounds are determined using an electrophoretic technique (see Section 5.5.1), or any of the methods disclosed in Section 5.5 *infra*. In another embodiment, the complexed target nucleic acid does not need to be separated from the free target nucleic acid if a technique (*i.e.*, spectrometry) that differentiates between bound and unbound target nucleic acids is used.

The methods for identifying small molecules bound to labeled nucleic acid will vary with the type of label on the target nucleic acid. For example, if a target RNA is labeled with a visible or fluorescent dye, the target RNA complexes are preferably identified using a chromatographic technique that separates bound from free target by an electrophoretic or size differential technique using individual reactions. The reactions corresponding to changes in the migration of the complexed RNA can be cross-referenced to the small molecule compound(s) added to said reaction. Alternatively, complexed target RNA can be screened *en masse* and then separated from free target RNA using an electrophoretic or size differential technique, the resultant complexed target is then analyzed using a mass spectrometric technique. In this fashion the bound small molecule can be identified on the basis of its molecular weight. In this reaction *a priori* knowledge of the exact molecular weights of all compounds within the library is known. In another embodiment, the test compounds bound to the target nucleic acid may not require separation from the unbound target nucleic acid if a technique such as, but not limited to, spectrometry is used.

### 5.5. Separation Methods for Screening Test Compounds

Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids. Methods that can be utilized for the physical separation of complexed target RNA from unbound target RNA include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships (“SAR”) by NMR spectroscopy, size exclusion chromatography, affinity chromatography,

and nanoparticle aggregation.

#### 5.5.1. Electrophoresis

5 Methods for separation of the complex of a target RNA bound to a test compound from the unbound RNA comprises any method of electrophoretic separation, including but not limited to, denaturing and non-denaturing polyacrylamide gel electrophoresis, urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, agarose gel electrophoresis, and capillary electrophoresis.

10 In a preferred embodiment, an automated electrophoretic system comprising a capillary cartridge having a plurality of capillary tubes is used for high-throughput screening of test compounds bound to target RNA. Such an apparatus for performing automated capillary gel electrophoresis is disclosed in U.S. Patent Nos. 5,885,430; 15 5,916,428; 6,027,627; and 6,063,251, the disclosures of which are incorporated by reference in their entireties.

15 The device disclosed in U.S. Patent No. 5,885,430, which is incorporated by reference in its entirety, allows one to simultaneously introduce samples into a plurality of capillary tubes directly from microtiter trays having a standard size. U.S. Patent No. 5,885,430 discloses a disposable capillary cartridge which can be cleaned between 20 electrophoresis runs, the cartridge having a plurality of capillary tubes. A first end of each capillary tube is retained in a mounting plate, the first ends collectively forming an array in the mounting plate. The spacing between the first ends corresponds to the spacing between the centers of the wells of a microtiter tray having a standard size. Thus, the first ends of the capillary tubes can simultaneously be dipped into the samples present in the tray's wells. 25 The cartridge is provided with a second mounting plate in which the second ends of the capillary tubes are retained. The second ends of the capillary tubes are arranged in an array which corresponds to the wells in the microtiter tray, which allows for each capillary tube to be isolated from its neighbors and therefore free from cross-contamination, as each end is dipped into an individual well.

30 35 Plate holes may be provided in each mounting plate and the capillary tubes inserted through these plate holes. In such a case, the plate holes are sealed airtight so that the side of the mounting plate having the exposed capillary ends can be pressurized. Application of a positive pressure in the vicinity of the capillary openings in this mounting plate allows for the introduction of air and fluids during electrophoretic operations and also can be used to force out gel and other materials from the capillary tubes during

reconditioning. The capillary tubes may be protected from damage using a needle comprising a cannula and/or plastic tubes, and the like when they are placed in these plate holes. When metallic cannula or the like are used, they can serve as electrical contacts for current flow during electrophoresis. In the presence of a second mounting plate, the second mounting plate is provided with plate holes through which the second ends of the capillary tubes project. In this instance, the second mounting plate serves as a pressure containment member of a pressure cell and the second ends of the capillary tubes communicate with an internal cavity of the pressure cell. The pressure cell is also formed with an inlet and an outlet. Gels, buffer solutions, cleaning agents, and the like may be introduced into the internal cavity through the inlet, and each of these can simultaneously enter the second ends of the capillaries.

In another preferred embodiment, the automated electrophoretic system can comprise a chip system consisting of complex designs of interconnected channels that perform and analyze enzyme reactions using part of a channel design as a tiny, continuously operating electrophoresis material, where reactions with one sample are going on in one area of the chip while electrophoretic separation of the products of another sample is taking place in a different part of the chip. Such a system is disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of which are incorporated by reference in their entireties.

The system disclosed in U.S. Patent No. 5,699,157, which is hereby incorporated by reference in its entirety, provides for a microfluidic system for high-speed electrophoretic analysis of subject materials for applications in the fields of chemistry, 25 biochemistry, biotechnology, molecular biology and numerous other areas. The system has a channel in a substrate, a light source and a photoreceptor. The channel holds subject materials in solution in an electric field so that the materials move through the channel and separate into bands according to species. The light source excites fluorescent light in the species bands and the photoreceptor is arranged to receive the fluorescent light from the bands. The system further has a means for masking the channel so that the photoreceptor can receive the fluorescent light only at periodically spaced regions along the channel. The system also has an unit connected to analyze the modulation frequencies of light intensity received by the photoreceptor so that velocities of the bands along the channel are 30 determined, which allows the materials to be analyzed.

35 The system disclosed in U.S. Patent No. 5,699,157 also provides for a

method of performing high-speed electrophoretic analysis of subject materials, which comprises the steps of holding the subject materials in solution in a channel of a microfluidic system; subjecting the materials to an electric field so that the subject materials move through the channel and separate into species bands; directing light toward the channel; receiving light from periodically spaced regions along the channel simultaneously; and analyzing the frequencies of light intensity of the received light so that velocities of the bands along the channel can be determined for analysis of said materials. The determination of the velocity of a species band determines the electrophoretic mobility of the species and its identification.

5 U.S. Patent No. 5,842,787, which is hereby incorporated by reference in its entirety, is generally directed to devices and systems employ channels having, at least in part, depths that are varied over those which have been previously described (such as the device disclosed in U.S. Patent No. 5,699,157), wherein said channel depths provide numerous beneficial and unexpected results such as but not limited to, a reduction in sample perturbation, reduced non-specific sample mixture by diffusion, and increased resolution.

10 In another embodiment, the electrophoretic method of separation comprises polyacrylamide gel electrophoresis. In a preferred embodiment, the polyacrylamide gel electrophoresis is non-denaturing, so as to differentiate the mobilities of the target RNA bound to a test compound from free target RNA. If the polyacrylamide gel electrophoresis is denaturing, then the target RNA: test compound complex must be cross-linked prior to electrophoresis to prevent the disassociation of the target RNA from the test compound during electrophoresis. Such techniques are well known to one of skill in the art.

15 In one embodiment of the method, the binding of test compounds to target nucleic acid can be detected, preferably in an automated fashion, by gel electrophoretic analysis of interference footprinting. RNA can be degraded at specific base sites by enzymatic methods such as ribonucleases A, U<sub>2</sub>, CL<sub>3</sub>, T<sub>1</sub>, Phy M, and *B. cereus* or chemical methods such as diethylpyrocarbonate, sodium hydroxide, hydrazine, piperidine formate, dimethyl sulfate,

20 [2,12-dimethyl-3,7,11,17-tetraazacyclo[11.3.1]heptadeca-1(17),2,11,13,15-pentaenato] nickel(II) (NiCR), cobalt(II)chloride, or iron(II) ethylenediaminetetraacetate (Fe-EDTA) as described for example in Zheng *et al.*, 1999, Biochem. 37:2207-2214; Latham & Cech, 1989, Science 245:276-282; and Sambrook *et al.*, 2001, in Molecular Cloning: A Laboratory Manual, pp 12.61-12.73, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties. The

specific pattern of cleavage sites is determined by the accessibility of particular bases to the reagent employed to initiate cleavage and, as such, is therefore determined by the three-dimensional structure of the RNA.

5 The interaction of small molecules with a target nucleic acid can change the accessibility of bases to these cleavage reagents both by causing conformational changes in the target nucleic acid or by covering a base at the binding interface. When a test compound binds to the nucleic acid and changes the accessibility of bases to cleavage reagents, the observed cleavage pattern will change. This method can be used to identify  
10 and characterize the binding of small molecules to RNA as described, for example, by Prudent *et al.*, 1995, J. Am. Chem. Soc. 117:10145-10146 and Mei *et al.*, 1998, Biochem. 37:14204-14212.

15 In the preferred embodiment of this technique, the detectably labeled target nucleic acid is incubated with an individual test compound and then subjected to treatment with a cleavage reagent, either enzymatic or chemical. The reaction mixture can be  
20 preferably be examined directly, or treated further to isolate and concentrate the nucleic acid. The fragments produced are separated by electrophoresis and the pattern of cleavage can be compared to a cleavage reaction performed in the absence of test compound. A change in the cleavage pattern directly indicates that the test compound binds to the target nucleic acid. Multiple test compounds can be examined both in parallel and serially.

25 Other embodiments of electrophoretic separation include, but are not limited to urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, and agarose gel electrophoresis.

25

### 5.5.2. Fluorescence Spectroscopy

30 In a preferred embodiment, fluorescence polarization spectroscopy, an optical detection method that can differentiate the proportion of a fluorescent molecule that is either bound or unbound in solution (e.g., the labeled target nucleic acid of the present invention), can be used to read reaction results without electrophoretic separation of the samples. Fluorescence polarization spectroscopy can be used to read the reaction results in the chip system disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of  
35 which are incorporated by reference in their entireties. The application of fluorescence

polarization spectroscopy to the chip system disclosed in the U.S. Patents listed *supra* is fast, efficient, and well-adapted for high-throughput screening.

In another embodiment, a compound that has an affinity for the target nucleic acid of interest can be labeled with a fluorophore to screen for test compounds that bind to the target nucleic acid. For example, a pyrene-containing aminoglycoside analog was used to accurately monitor antagonist binding to a prokaryotic 16S rRNA A site (which comprises the natural target for aminoglycoside antibiotics) in a screen using a fluorescence quenching technique in a 96-well plate format (Hamasaki & Rando, 1998, *Anal. Biochem.* 261(2):183-90).

In another embodiment, fluorescence resonance energy transfer (FRET) can be used to screen for test compounds that bind to the target nucleic acid. FRET, a characteristic change in fluorescence, occurs when two fluorophores with overlapping emission and excitation wavelength bands are held together in close proximity, such as by a binding event. In the preferred embodiment, the fluorophore on the target nucleic acid and the fluorophore on the test compounds will have overlapping excitation and emission spectra such that one fluorophore (the donor) transfers its emission energy to excite the other fluorophore (the acceptor). The acceptor preferably emits light of a different wavelength upon relaxing to the ground state, or relaxes non-radiatively to quench fluorescence. FRET is very sensitive to the distance between the two fluorophores, and allows measurement of molecular distances less than 10 nm. For example, U.S. Patent 6,337,183 to Arenas *et al.*, which is incorporated by reference in its entirety, describes a screen for compounds that bind RNA that uses FRET to measure the effect of test compounds on the stability of a target RNA molecule where the target RNA is labeled with both fluorescent acceptor and donor molecules and the distance between the two fluorophores as determined by FRET provides a measure of the folded structure of the RNA. Matsumoto *et al.* (2000, *Bioorg. Med. Chem. Lett.* 10:1857-1861) describe a system where a peptide that binds to HIV-1 TAR RNA is labeled on one end with a fluorescein fluorophore and a tetramethylrhodamine on the other end. The conformational change of the peptide upon binding to the RNA provided a FRET signal to screen for compounds that bound to the TAR RNA.

In the preferred embodiment, both the target nucleic acid and a compound that has an affinity for the target nucleic acid of interest are labeled with fluorophores with overlapping emission and excitation spectra (donor and acceptor), including but not limited to fluorescein and derivatives, rhodamine and derivatives, cyanine dyes and derivatives, bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, pyrene, nanoparticles, or

non-fluorescent quenching molecules. Binding of a labeled test compound to the target nucleic acid can be identified by the change in observable fluorescence as a result of FRET.

If the target nucleic acid is labeled with the donor fluorophore, then the test compounds is labeled with the acceptor fluorophore. Conversely, if the target nucleic acid is labeled with the acceptor fluorophore, then the test compounds is labeled with the donor fluorophore. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions. The fluorophore on the target nucleic acid must be in close proximity to the binding site of the test compounds, but should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site.

In yet another embodiment, homogeneous time-resolved fluorescence (“HTRF”) techniques based on time-resolved energy transfer from lanthanide ion complexes to a suitable acceptor species can be adapted for high-throughput screening for inhibitors of RNA-protein complexes (Hemmilä, 1999, *J. Biomol. Screening* 4:303-307; Mathis, 1999, *J. Biomol. Screening* 4:309-313). HTRF is similar to fluorescence resonance energy transfer using conventional organic dye pairs, but has several advantages, such as increased sensitivity and efficiency, and background elimination (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356).

Fluorescence spectroscopy has traditionally been used to characterize DNA-protein and protein-protein interactions, but fluorescence spectroscopy has not been widely used to characterize RNA-protein interactions because of an interfering absorption of RNA nucleotides with the intrinsic tryptophan fluorescence of proteins (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356.). However, fluorescence spectroscopy has been used in studying the single tryptophan residue within the arginine-rich RNA-binding domain of Rev protein and its interaction with the RRE in a time-resolved fluorescence study (Kwon & Carson, 1998, *Anal. Biochem.* 264:133-140). Thus, in this invention, fluorescence spectroscopy is less preferred if the test compounds or peptides or proteins possess intrinsic tryptophan fluorescence. However, fluorescence spectroscopy can be used for test compounds that do not possess intrinsic fluorescence.

#### **5.5.3. Surface Plasmon Resonance (“SPR”)**

Surface plasmon resonance (SPR) can be used for determining kinetic rate constants and equilibrium constants for macromolecular interactions by following the

association project in "real time" (Schuck, 1997, *Annu. Rev. Biophys. Biomol. Struct.* 26:541-566).

The principle of SPR is summarized by Xavier *et al.* (*Trends Biotechnol.*, 2000, 18(8):349-356) as follows. Total internal reflection occurs at the boundary between two substances of different refractive index. The incident light's electromagnetic field penetrates beyond the interface as an evanescent wave, which extends a few hundred nanometers beyond the surface into the medium. Insertion of a thin gold foil at the interface produced SPR owing to the absorption of the energy from the evanescent wave by free electron clouds of the metal (plasmons). As a result of this absorbance, there is a drop in the intensity of the reflected light at a particular angle of incidence. The evanescent wave profile depends exquisitely on the refractive index of the medium it probes. Thus, the angle at which absorption occurs is very sensitive to the refractive changes in the external medium. All proteins and nucleic acids are known to change the refractive index of water by a similar amount per unit mass, irrespective of their amino acid or nucleotide composition (the refractive index change is different for proteins and nucleic acids). When the protein or nucleic acid content of the layer at the sensor changes, the refractive index also changes. Typically, one member of a complex is immobilized in a dextran layer and then the other member is introduced into the solution, either in a flow cell (Biacore AB, Uppsala, Sweden) or a stirred cuvette (Affinity Sensors, Santa Fe, New Mexico). It has been determined that there is a linear correlation between the surface concentration of protein or nucleic acid and the shift in resonance angle, which can be used to quantitate kinetic rate constants and/or the equilibrium constants.

In the present invention, the target RNA may be immobilized to the sensor surface through a streptavidin-biotin linkage, the linkage of which is disclosed by Crouch *et al.* (*Methods Mol. Biol.*, 1999, 118:143-160). The RNA is biotinylated either during synthesis or post-synthetically via the conversion of the 3' terminal ribonucleoside of the RNA into a reactive free amino group or using a T7 polymerase incorporated guanosine monophosphorothioate at the 5' end. SPR has been used to determine the stoichiometry and affinity of the interaction between the HIV Rev protein and the RRE (Van Ryk & Venkatesan, 1999, *J. Biol. Chem.* 274:17452-17463) and the aminoglycoside antibiotics with RRE and a model RNA derived from the 16S ribosomal A site, respectively (Hendrix *et al.*, 1997, *J. Am. Chem. Soc.* 119:3641-3648; Wong *et al.*, 1998, *Chem. Biol.* 5:397-406).

In one embodiment of the present invention, the target nucleic acid can be immobilized to a sensor surface (e.g., by a streptavidin-biotin linkage) and SPR can be used

to (a) determine whether the target RNA binds a test compound and (b) further characterize the binding of the target nucleic acids of the present invention to a test compound.

#### 5.5.4. Mass Spectrometry

An automated method for analyzing mass spectrometer data which can analyze complex mixtures containing many thousands of components and can correct for background noise, multiply charged peaks and atomic isotope peaks is described in U.S. Patent No. 6,147,344, which is hereby incorporated by reference in its entirety. The system disclosed in U.S. Patent No. 6,147,344 is a method for analyzing mass spectrometer data in which a control sample measurement is performed providing a background noise check. The peak height and width values at each m/z ratio as a function of time are stored in a memory. A mass spectrometer operation on a material to be analyzed is performed and the peak height and width values at each m/z ratio versus time are stored in a second memory location. The mass spectrometer operation on the material to be analyzed is repeated a fixed number of times and the stored control sample values at each m/z ratio level at each time increment are subtracted from each corresponding one from the operational runs, thus producing a difference value at each mass ratio for each of the multiple runs at each time increment. If the MS value minus the background noise does not exceed a preset value, the m/z ratio data point is not recorded, thus eliminating background noise, chemical noise and false positive peaks from the mass spectrometer data. The stored data for each of the multiple runs is then compared to a predetermined value at each m/z ratio and the resultant series of peaks, which are now determined to be above the background, is stored in the m/z points in which the peaks are of significance.

25 One possibility for the utilization of mass spectrometry in high throughput screening is the integration of SPR with mass spectrometry. Approaches that have been tried are direct analysis of the analyte retained on the sensor chip and mass spectrometry with the eluted analyte (Sonksen *et al.*, 1998, *Anal. Chem.* 70:2731-2736; Nelson & Krone, 1999, *J. Mol. Recog.* 12:77-93). Further developments, especially in the interfacing of the sensor chip with the mass spectrometer and in reusing the sensor chip, are required to make SPR combined with mass spectroscopy a high-throughput method for biomolecular interaction analysis and the screening of targets for small molecule inhibitors (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356).

30

In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by any of the mass spectrometry processes described *supra*. Furthermore, mass spectrometry can also be used to elucidate

the structure of the test compound.

#### **5.5.5. Scintillation Proximity Assay (“SPA”)**

Scintillation Proximity Assay (“SPA”) is a method that can be used for screening small molecules that bind to the target RNAs. SPA would involve radiolabeling either the target RNA or the test compound and then quantitating its binding to the other member to a bead or a surface impregnated with a scintillant (Cook, 1996, *Drug Discov. Today* 1:287-294). Currently, fluorescence-based techniques are preferred for high-throughput screening (Pope *et al.*, 1999, *Drug Discov. Today* 4:350-362).

10 Screening for small molecules that inhibit Tat peptide: TAR RNA interaction has been performed with SPA, and inhibitors of the interaction were isolated and characterized (Mei *et al.*, 1997, *Bioorg. Med. Chem.* 5:1173-1184; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212). A similar approach can be used to identify small 15 molecules that directly bind to a preselected target RNA element in accordance with the invention can be utilized.

SPA can be adapted to high throughput screening by the availability of microplates, wherein the scintillant is directly incorporated into the plastic of the microtiter wells (Nakayama *et al.*, 1998, *J. Biomol. Screening* 3:43-48). Thus, one embodiment of the 20 present invention comprises (a) labeling of the target nucleic acid with a radioactive or fluorescent label; (b) contacted the labeled nucleic acid with test compounds, wherein each test compound is in a microtiter well coated with scintillant and is tethered to the microtiter well; and (c) identifying and quantifying the test compounds bound to the target nucleic acid with SPA, wherein the test compound is identified by virtue of its location in the 25 microplate.

#### **5.5.6. Structure-Activity Relationships (“SAR”) by NMR Spectroscopy**

NMR spectroscopy is a valuable technique for identifying complexed target nucleic acids by qualitatively determining changes in chemical shift, specifically from 30 distances measured using relaxation effects, and NMR-based approaches have been used in the identification of small molecule binders of protein drug targets (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). The determination of structure-activity relationships (“SAR”) by NMR is the first method for NMR described in which small molecules that bind adjacent subsites are identified by two-dimentional <sup>1</sup>H-<sup>15</sup>N spectra of the target protein 35 (Shuker *et al.*, 1996, *Science* 274:1531-1534). The signal from the bound molecule is monitored by employing line broadening, transferred NOEs and pulsed field gradient

diffusion measurements (Moore, 1999, *Curr. Opin. Biotechnol.* 10:54-58). A strategy for lead generation by NMR using a library of small molecules has been recently described (Fejzo *et al.*, 1999, *Chem. Biol.* 6:755-769).

5 In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by SAR by NMR. Furthermore, SAR by NMR can also be used to elucidate the structure of the test compound.

#### **5.5.7. Size Exclusion Chromatography**

10 In another embodiment of the present invention, size-exclusion chromatography is used to purify test compounds that are bound to a target nucleic acid from a complex mixture of compounds. Size-exclusion chromatography separates molecules based on their size and uses gel-based media comprised of beads with specific size distributions. When applied to a column, this media settles into a tightly packed matrix and forms a complex array of pores. Separation is accomplished by the inclusion or exclusion 15 of molecules by these pores based on molecular size. Small molecules are included into the pores and, consequently, their migration through the matrix is retarded due to the added distance they must travel before elution. Large molecules are excluded from the pores and migrate with the void volume when applied to the matrix. In the present invention, a target nucleic acid is incubated with a mixture of test compounds while free in solution and 20 allowed to reach equilibrium. When applied to a size exclusion column, test compounds free in solution are retained by the column, and test compounds bound to the target nucleic acid are passed through the column. In a preferred embodiment, spin columns commonly used for "desalting" of nucleic acids will be employed to separate bound from unbound test 25 compounds (e.g., Bio-Spin columns manufactured by BIO-RAD). In another embodiment, the size exclusion matrix is packed into multiwell plates to allow high throughput separation of mixtures (e.g., PLASMID 96-well SEC plates manufactured by Millipore).

#### **5.5.8. Affinity Chromatography**

30 In one embodiment of the present invention, affinity capture is used to purify test compounds that are bound to a target nucleic acid labeled with an affinity tag from a complex mixture of compounds. To accomplish this, a target nucleic acid labeled with an affinity tag is incubated with a mixture of test compounds while free in solution and then captured to a solid support once equilibrium has been established; alternatively, target 35 nucleic acids labeled with an affinity tag can be captured to a solid support first and then allowed to reach equilibrium with a mixture of test compounds.

The solid support is typically comprised of, but not limited to, cross-linked agarose beads that are coupled with a ligand for the affinity tag. Alternatively, the solid support may be a glass, silicon, metal, or carbon, plastic (polystyrene, polypropylene) surface with or without a self-assembled monolayer (SAM) either with a covalently attached ligand for the affinity tag, or with inherent affinity for the tag on the target nucleic acid.

Once the complex between the target nucleic acid and test compound has reached equilibrium and has been captured, one skilled in the art will appreciate that the retention of bound compounds and removal of unbound compounds is facilitated by washing the solid support with large excesses of binding reaction buffer. Furthermore, retention of high affinity compounds and removal of low affinity compounds can be accomplished by a number of means that increase the stringency of washing; these means include, but are not limited to, increasing the number and duration of washes, raising the salt concentration of the wash buffer, addition of detergent or surfactant to the wash buffer, and addition of non-specific competitor to the wash buffer.

In one embodiment, the test compounds themselves are detectably labeled with fluorescent dyes, radioactive isotopes, or nanoparticles. When the test compounds are applied to the captured target nucleic acid in a spatially addressed fashion (e.g., in separate wells of a 96-well microplate), binding between the test compounds and the target nucleic acid can be determined by the presence of the detectable label on the test compound using fluorescence.

Following the removal of unbound compounds, bound compounds with high affinity for the target nucleic acid can be eluted from the immobilized target nucleic acids and analyzed. The elution of test compounds can be accomplished by any means that break the non-covalent interactions between the target nucleic acid and compound. Means for elution include, but are not limited to, changing the pH, changing the salt concentration, the application of organic solvents, and the application of molecules that compete with the bound ligand. In a preferred embodiment, the means employed for elution will release the compound from the target RNA, but will not effect the interaction between the affinity tag and the solid support, thereby achieving selective elution of test compound. Moreover, a preferred embodiment will employ an elution buffer that is volatile to allow for subsequent concentration by lyophilization of the eluted compound (e.g., 0 M to 5 M ammonium acetate).

### **5.5.9. Nanoparticle Aggregation**

In one embodiment of the present invention, both the target nucleic acid and the test compounds are labeled with nanoparticles. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to  $\text{Ag}_2\text{S}$ ,  $\text{ZnS}$ ,  $\text{CdS}$ ,  $\text{CdTe}$ ,  $\text{Au}$ , or  $\text{TiO}_2$ . Methods for the attachment of nucleic acids and small molecules to nanoparticles are well known to one of skill in the art (reviewed in Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40:4129-4158. The references cited therein are hereby incorporated by reference in their entireties). In particular, if multiple copies of the target nucleic acid are attached to a single nanoparticle and multiple copies of a test compound are attached to another nanoparticle, then interaction between the test compound and target nucleic acid will induce aggregation of nanoparticles as described, for example, by Mitchel *et al.* 1999, *J. Am. Chem. Soc.* 121:8122-8123. The aggregate can be detected by changes in absorbance or fluorescence spectra and physically separated from the unbound components through filtration or centrifugation.

### **5.6. Methods for Identifying or Characterizing the Test Compounds Bound to the Target Nucleic Acids**

If the library comprises arrays or microarrays of test compounds, wherein each test compound has an address or identifier, the test compound can be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays.

If the library is a peptide or nucleic acid library, the sequence of the test compound can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

A number of physico-chemical techniques can be used for the *de novo* characterization of test compounds bound to the target.

#### **5.6.1. Mass Spectrometry**

Mass spectrometry (*e.g.*, electrospray ionization (“ESI”) and matrix-assisted laser desorption-ionization (“MALDI”), Fourier-transform ion cyclotron resonance (“FT-ICR”)) can be used both for high-throughput screening of test compounds that bind to a target RNA and elucidating the structure of the test compound. Thus, one example of mass spectroscopy is that separation of a bound and unbound complex and test compound structure elucidation can be carried out in a single step.

MALDI uses a pulsed laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovulj *et al.*, 1997, *J. Biol. Chem.* 272:32084-32091).  
5 However, covalent cross-linking between the target nucleic acid and the test compound is required for detection, since a non-covalently bound complex may dissociate during the MALDI process.

ESI mass spectrometry (“ESI-MS”) has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates 10 molecular ions with little to no fragmentation (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). ESI-MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, 1997, *Anal. Chem.* 69:5130-5135).

Fourier-transform ion cyclotron resonance (“FT-ICR”) mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate 15 mass assignments (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, 1999, *Anal. Chem.* 71:3436-3440; Griffey *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96:10129-10133). As true for all of the mass spectrometry methods discussed herein, FT-ICR does not require labeling of the target RNA or a test 20 compound.

An advantage of mass spectroscopy is not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the 25 discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

#### 5.6.2. NMR Spectroscopy

As described above, NMR spectroscopy is a technique for identifying 30 binding sites in target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects. Examples of NMR that can be used for the invention include, but are not limited to, one-dimensional NMR, two-dimensional NMR, correlation spectroscopy (“COSY”), and nuclear Overhauser effect (“NOE”) spectroscopy. Such methods of structure determination of test compounds are well known to one of skill in the art.

Similar to mass spectroscopy, an advantage of NMR is the not only the 35 elucidation of the structure of the test compound, but also the determination of the structure

of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

5

### 5.6.3. Vibrational Spectroscopy

Vibrational spectroscopy (e.g. infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of the test compound on the isolated bead.

10

Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the test compound as a result of excitation of vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

15

Infrared spectra can be measured in a scanning mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed-frequency infrared light source, by the test compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument). In a preferred embodiment, infrared spectra are measured in a pulsed mode (FT-IR) where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the test compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

20

Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam. The incident monochromatic light beam, usually a single laser frequency, is not truly absorbed by the test compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule. While some vibrational transitions are observable in both infrared and Raman spectrometry, most are observable only with one or

the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

5 Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby incorporated by reference.

10 Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Reffner *et al.*, which is hereby incorporated by reference in its entirety. An instrument that simultaneously performs a microscopic infrared and microscopic Raman analysis on a 15 sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby incorporated by reference in its entirety.

20 In the preferred embodiment, test compounds can be identified by matching the IR or Raman spectra of a test compound to a dataset of vibrational (IR or Raman) spectra previously acquired for each compound in the combinatorial library. By this method, the spectra of compounds with known structure are recorded so that comparison with these spectra can identify compounds again when isolated from RNA binding experiments.

### **5.7. Secondary Biological Screens**

25 The test compounds identified in the binding assay (for convenience referred to herein as a “lead” compound) can be tested for biological activity using host cells containing or engineered to contain the target RNA element coupled to a functional readout system. For example, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene. In this example, the 30 lead compounds are assayed in the presence or absence of the target RNA. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound.

35 In one embodiment, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene, such as, but not limited to,  $\beta$ -galactosidase, green fluorescent protein, red fluorescent protein, luciferase, chloramphenicol acetyltransferase, alkaline phosphatase, and  $\beta$ -

lactamase. In a preferred embodiment, a cDNA encoding the target element is fused upstream to a reporter gene wherein translation of the reporter gene is repressed upon binding of the lead compound to the target RNA. In other words, the steric hindrance caused by the binding of the lead compound to the target RNA repressed the translation of the reporter gene. This method, termed the translational repression assay procedure ("TRAP") has been demonstrated in *E. coli* and *S. cerevisiae* (Jain & Belasco, 1996, *Cell* 87(1):115-25; Huang & Schreiber, 1997, *Proc. Natl. Acad. Sci. USA* 94:13396-13401).

In another embodiment, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound. For example, the target RNA may be overexpressed in a cell in which the target RNA is endogenously expressed. Where the target RNA controls expression of a gene product involved in cell growth or viability, the *in vivo* effect of the lead compound can be assayed by measuring the cell growth or viability of the target cell. Alternatively, a reporter gene can also be fused downstream of the target RNA sequence and the effect of the lead compound on reporter gene expression can be assayed.

Alternatively, the lead compounds identified in the binding assay can be tested for biological activity using animal models for a disease, condition, or syndrome of interest. These include animals engineered to contain the target RNA element coupled to a functional readout system, such as a transgenic mouse. Animal model systems can also be used to demonstrate safety and efficacy.

Compounds displaying the desired biological activity can be considered to be lead compounds, and will be used in the design of congeners or analogs possessing useful pharmacological activity and physiological profiles. Following the identification of a lead compound, molecular modeling techniques can be employed, which have proven to be useful in conjunction with synthetic efforts, to design variants of the lead that can be more effective. These applications may include, but are not limited to, Pharmacophore Modeling (*cf.* Lamothe, *et al.* 1997, *J. Med. Chem.* 40: 3542; Mottola *et al.* 1996, *J. Med. Chem.* 39: 285; Beusen *et al.* 1995, *Biopolymers* 36: 181; P. Fossa *et al.* 1998, *Comput. Aided Mol. Des.* 12: 361), QSAR development (*cf.* Siddiqui *et al.* 1999, *J. Med. Chem.* 42: 4122; Barreca *et al.* 1999 *Bioorg. Med. Chem.* 7: 2283; Kroemer *et al.* 1995, *J. Med. Chem.* 38: 4917; Schaal *et al.* 2001, *J. Med. Chem.* 44: 155; Buolamwini & Assefa 2002, *J. Mol. Chem.* 45: 84), Virtual docking and screening/scoring (*cf.* Anzini *et al.* 2001, *J. Med. Chem.* 44: 1134; Faaland *et al.* 2000, *Biochem. Cell. Biol.* 78: 415; Silvestri *et al.* 2000, *Bioorg. Med. Chem.* 8: 2305; J. Lee *et al.* 2001, *Bioorg. Med. Chem.* 9: 19), and Structure Prediction using RNA structural programs including, but not limited to mFold (as described

5 by Zuker *et al.* Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology pp. 11-43, J. Barciszewski & B.F.C. Clark, eds. (NATO ASI Series, Kluwer Academic Publishers, 1999) and Mathews *et al.* 1999 J. Mol. Biol. 288: 911-940); RNAMotif (Macke *et al.* 2001, Nucleic Acids Res. 29: 4724-4735; and the Vienna RNA package (Hofacker *et al.* 1994, Monatsh. Chem. 125: 167-188).

10 Further examples of the application of such techniques can be found in several review articles, such as Rotivinen *et al.*, 1988, Acta Pharmaceutical Fennica 97:159-166; Ripka, 1998, New Scientist 54-57; McKinlay & Rossmann, 1989, Annu. Rev. Pharmacol. Toxicol. 29:111-122; Perry & Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, Proc. R. Soc. Lond. 236:125-140 and 141-162; Askew *et al.*, 1989, J. Am. Chem. Soc. 111:1082-1090. Molecular modeling tools employed may include those from Tripos, Inc., 15 St. Louis, Missouri (e.g., Sybyl/UNITY, CONCORD, DiverseSolutions), Accelrys, San Diego, California (e.g., Catalyst, Wisconsin Package {BLAST, etc.}), Schrodinger, Portland, Oregon (e.g., QikProp, QikFit, Jaguar) or other such vendors as BioDesign, Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario, Canada), and may include privately designed and/or "academic" 20 software (e.g. RNAMotif, mFOLD). These application suites and programs include tools for the atomistic construction and analysis of structural models for drug-like molecules, proteins, and DNA or RNA and their potential interactions. They also provide for the calculation of important physical properties, such as solubility estimates, permeability metrics, and empirical measures of molecular "druggability" (e.g., Lipinski "Rule of 5" as 25 described by Lipinski *et al.* 1997, Adv. Drug Delivery Rev. 23: 3-25). Most importantly, they provide appropriate metrics and statistical modeling power (such as the patented CoMFA technology in Sybyl as described in US Patents 6,240,374 and 6,185,506) to develop Quantitative Structural Activity Relationships (QSARs) which are used to guide the synthesis of more efficacious clinical development candidates while improving 30 desirable physical properties, as determined by results from the aforementioned secondary screening protocols.

#### **5.8. Use of Identified Compounds That Bind RNA to Treat/Prevent Disease**

35 Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from a disease whose progression is

associated with a target RNA:host cell factor interaction *in vivo*. In certain embodiments, such compounds or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*.

5 In one embodiment, "treatment" or "treating" refers to an amelioration of a disease, or at least one discernible symptom thereof. In another embodiment, "treatment" or "treating" refers to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In yet another embodiment, "treatment" or "treating" refers to inhibiting the progression of a disease, either physically, *e.g.*, stabilization of a 10 discernible symptom, physiologically, *e.g.*, stabilization of a physical parameter, or both. In yet another embodiment, "treatment" or "treating" refers to delaying the onset of a disease.

10 In certain embodiments, the compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction 15 *in vivo*. As used herein, "prevention" or "preventing" refers to a reduction of the risk of acquiring a disease. In one embodiment, the compound or a pharmaceutically acceptable salt thereof is administered as a preventative measure to a patient. According to this embodiment, the patient can have a genetic predisposition to a disease, such as a family 20 history of the disease, or a non-genetic predisposition to the disease. Accordingly, the compound and pharmaceutically acceptable salts thereof can be used for the treatment of one manifestation of a disease and prevention of another.

25 When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that optionally comprises a pharmaceutically acceptable vehicle. The composition can be administered orally, or by any other convenient route, for example, by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery 30 systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

35 Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to

the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

5 In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

10 15 In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

15 20 Pulmonary administration can also be employed, *e.g.*, by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

20 25 In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, *Science* 249:1527-1533; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

25 30 35 In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, *e.g.*, Goodson, in *Medical Applications of Controlled Release*, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, *Science* 249:1527-1533) may be used. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald *et al.*, 1980, *Surgery* 88:507; Saudek *et al.*, 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (see *Medical Applications of Controlled Release*, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); *Controlled Drug Bioavailability, Drug Product Design and Performance*, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, 1983, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; see also Levy *et al.*, 1985, *Science*

228:190; During *et al.*, 1989, Ann. Neurol. 25:351; Howard *et al.*, 1989, J. Neurosurg. 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

5 Compositions comprising the compound or a pharmaceutically acceptable salt thereof ("compound compositions") can additionally comprise a suitable amount of a pharmaceutically acceptable vehicle so as to provide the form for proper administration to the patient.

10 In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, mammals, and more particularly in humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is administered. Such pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to 15 a patient, the pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, 20 silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Compound compositions, if 25 desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents.

30 Compound compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see *e.g.*, U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in Remington's Pharmaceutical Sciences, Alfonso R. Gennaro, ed., 35 Mack Publishing Co. Easton, PA, 19th ed., 1995, pp. 1447 to 1676, incorporated herein by reference.

In a preferred embodiment, the compound or a pharmaceutically acceptable salt thereof is formulated in accordance with routine procedures as a pharmaceutical composition adapted for oral administration to human beings. Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, 5 powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions may contain one or more agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable preparation. Moreover, where in tablet or pill form, the compositions can be coated to 10 delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained action over an extended period of time. Selectively permeable membranes surrounding an osmotically active driving compound are also suitable for orally administered compositions. In these later platforms, fluid from the environment 15 surrounding the capsule is imbibed by the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can include standard vehicles such as 20 mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. Such vehicles are preferably of pharmaceutical grade. Typically, compositions for intravenous administration comprise sterile isotonic aqueous buffer. 25 Where necessary, the compositions may also include a solubilizing agent.

In another embodiment, the compound or a pharmaceutically acceptable salt thereof 25 can be formulated for intravenous administration. Compositions for intravenous administration may optionally include a local anesthetic such as lignocaine to lessen pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free 30 concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the compound or a pharmaceutically acceptable salt thereof is to be administered by infusion, it can be dispensed, for example, with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the compound or a pharmaceutically acceptable salt thereof is administered by injection, an ampoule of sterile 35 water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The amount of a compound or a pharmaceutically acceptable salt thereof that will be effective in the treatment of a particular disease will depend on the nature of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The 5 precise dose to be employed will also depend on the route of administration, and the seriousness of the disease, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for oral administration are generally about 0.001 milligram to about 200 milligrams of a compound or a pharmaceutically acceptable salt thereof per kilogram body weight per day. In specific 10 preferred embodiments of the invention, the oral dose is about 0.01 milligram to about 100 milligrams per kilogram body weight per day, more preferably about 0.1 milligram to about 75 milligrams per kilogram body weight per day, more preferably about 0.5 milligram to 5 milligrams per kilogram body weight per day. The dosage amounts described herein refer 15 to total amounts administered; that is, if more than one compound is administered, or if a compound is administered with a therapeutic agent, then the preferred dosages correspond to the total amount administered. Oral compositions preferably contain about 10% to about 95% active ingredient by weight.

Suitable dosage ranges for intravenous (i.v.) administration are about 0.01 20 milligram to about 100 milligrams per kilogram body weight per day, about 0.1 milligram to about 35 milligrams per kilogram body weight per day, and about 1 milligram to about 10 milligrams per kilogram body weight per day. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight per day to about 1 mg/kg body weight per day. Suppositories generally contain about 0.01 milligram to about 50 25 milligrams of a compound of the invention per kilogram body weight per day and comprise active ingredient in the range of about 0.5% to about 10% by weight.

Recommended dosages for intradermal, intramuscular, intraperitoneal, 30 subcutaneous, epidural, sublingual, intracerebral, intravaginal, transdermal administration or administration by inhalation are in the range of about 0.001 milligram to about 200 milligrams per kilogram of body weight per day. Suitable doses for topical administration are in the range of about 0.001 milligram to about 1 milligram, depending on the area of administration. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems. Such animal models and systems are well known in the art.

35 The compound and pharmaceutically acceptable salts thereof are preferably assayed *in vitro* and *in vivo*, for the desired therapeutic or prophylactic activity, prior to use

in humans. For example, *in vitro* assays can be used to determine whether it is preferable to administer the compound, a pharmaceutically acceptable salt thereof, and/or another therapeutic agent. Animal model systems can be used to demonstrate safety and efficacy.

5 A variety of compounds can be used for treating or preventing diseases in mammals. Types of compounds include, but are not limited to, peptides, peptide analogs including peptides comprising non-natural amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as  $\alpha$ -amino phosphonic acids and  $\alpha$ -amino phosphinic acids, or amino acids having non-peptide linkages, nucleic acids, nucleic acid analogs such as phosphorothioates or peptide nucleic acids (“PNAs”), hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose.

10

## 6. EXAMPLE: THERAPEUTIC TARGETS

The therapeutic targets presented herein are by way of example, and the present invention is not to be limited by the targets described herein. The therapeutic targets presented herein as DNA sequences are understood by one of skill in the art that the sequences can be converted to RNA sequences.

## 6.1. Tumor Necrosis Factor Alpha (“TNF- $\alpha$ ”)

GenBank Accession # X01394:

1 gcagaggacc agctaagagg gagagaagca actacagacc cccctgaaa acaaccctca  
61 gacgccatat cccctgacaa gctgccaggc aggttcctt cctctcacat actgaccac  
121 ggctccaccc tctctccctt ggaaaggaca ccatgagcac tgaaagcatg atccgggacg  
181 tggagctggc cgaggaggcg ctccccaga agacaggggg gccccaggc tccaggcggt  
241 gcttgttctt cagcccttc tccttcttga tcgtggcagg cgccaccacg ctcttcgccc  
301 tgctgcactt tggagtgatc ggccccaga gggaaagagttt ccccagggac ctctcttcaa  
361 ttagccctctt ggcccaggca gtcagatcat ctctcgaac cccgagtgac aagcctgttag  
421 cccatgttgtt agcaaaccctt caagctgagg ggcagctcca gtggctgaac cgccgggcca  
481 atgccttcctt ggccaatggc gtggagctga gagataacca gctgggtggc ccatcagagg  
541 gcctgtacccatctactcc caggctcttcaaggccca aggctgcccc tccacccatg  
601 tgctcctcac ccacaccatc agccgcattcg ccgttcctta ccagaccaag gtcaaccctcc  
661 tctctgccccat caagagcccc tgcgcagaggg agacccaga gggggctgag gccaagccct  
721 ggtatgagcc cttttatctg ggaggggtctt tccagctgga gaagggtgac cgactcagcg  
781 cttagatcaa tcggcccgac tatctcgact ttgcgcgatc tggcgcaggc tactttggga

841 tcattgccct gtgaggagga cgaacatcca accttcccaa acgcctcccc tgccccaatc  
 901 cctttattac cccctccctc agacaccctc aaccctttct ggctaaaaaa gagaattggg  
 961 ggcttagggt cggaacccaa gcttagaact ttaagcaaca agaccaccac ttgaaacct  
 5 1021 gggattcagg aatgtgtggc ctgcacagtg aattgctggc aaccactaag aattcaaact  
 1081 gggccctcca gaactcactg gggcctacag ctttgcattcc tgacatctgg aatctggaga  
 1141 ccagggagcc ttgggtctg gccagaatgc tgcaggactt gagaagaccc cacctagaaa  
 1201 ttgacacaag tggaccttag gcctccctt ctccagatgt ttccagactt ctttgagaca  
 1261 cgaggcccag ccctcccat ggagccagct ccctctattt atgtttgcac ttgtgattat  
 1321 ttattattta ttattattt atttatttac agatgaatgt atttatttgg gagaccgggg  
 1381 tattctgggg gacccaatgt aggagctgcc ttggctcaga catgtttcc gtgaaaacgg  
 1441 agctgaacaa taggtgttc ccatgtgcc ccctggcctc tttgcattct ttgtgattatg  
 1501 tttttaaaaa tatttacatcg attaagtgtt ctaaacaatgt ctgattttggt gaccaactgt  
 1561 cactcattgc tgagcctctg ctccccaggg gagttgtgtc tggatcgcc ctactattca  
 15 1621 gtggcgagaa ataaagttt ctt (SEQ ID NO: 6)

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 152
- (2) 3' Untranslated Region - nts 852 - 1643

20 20 Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region  
 5' AUUUUUUUUUUUUUUUUUUA 3' (SEQ ID NO: 1)

25 25 **6.2. Granulocyte-macrophage Colony Stimulating Factor (“GM-CSF”)**  
 GenBank Accession # NM\_000758:

1 gctggaggat gtggctgcag agcctgctgc tctggcaca tggcctgc agcatctctg  
 61 cacccggcccg ctgcgccagc cccagcacgc agccctggga gcatgtaat gccatccagg  
 121 aggccggcgc ttcctgaac ctgagtagag acactgctgc tgagataat gaaacagtag  
 181 aagtcatctc agaaatgtt gaccicccagg agccgacccg cctacagacc cgcctggagc  
 241 tgtacaagca gggcctgcgg ggcagcctca ccaagctaa gggcccttg accatgtatgg  
 30 301 ccagccacta caagcagcac tgcctccaa ccccgaaac ttccctgtca acccagacta  
 361 tcaccttga aagttcaaa gagaacctga aggactttct gcttgcattt cccttgact  
 421 gctggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc  
 481 tctctcatga aacaagact agaaactcgat gatggtcattt ttggagggac caaggggtgg  
 541 gcccacagcca tggtggagt ggcctggacc tgcctgggc cacactgacc ctgatacagg

601 catggcagaa gaatgggaat atttatact gacagaaatc agtaatattt atatatttat  
 661 attttaaaa tatttatttta ttatttattt taagttcata ttccatattt attcaagatg  
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 7)

5

GenBank Accession # XM\_003751:

1 tctggaggat gtggctgcag agcctgctgc tctggcac tggcctgc agcatctgc  
 61 cacccgcggc ctgcggcggc cccagcacgc agccctggga gcatgtgaat gccatccagg  
 121 aggcccggcg tctcctgaac ctgagtagag acactgctgc tgagatgaat gaaacagtag  
 181 aagtcatctc agaaatgtt gacccctcagg agccgacccctg cctacagacc cgcctggc  
 241 tgtacaagca gggctgcgg ggcagcctca ccaagctaa gggcccttg accatgtatgg  
 301 ccagccacta caagcagcac tgccctccaa ccccgaaac ttccgtgca acccagacta  
 361 tcaccccttga aagttcaaa gagaacctga aggacttct gcttgcattc ccccttgact  
 421 gctgggagcc agtccaggag tgagacccggc cagatgaggc tggccaagcc ggggagctgc  
 481 tctctcatga aacaagagct agaaactcag gatggcattc ttggagggac caaggggtgg  
 541 gcccacggca tggggaggt ggcctggacc tgccctggc cacactgacc ctgatacagg  
 601 catggcagaa gaatgggaat atttatact gacagaaatc agtaatattt atatatttat  
 661 attttaaaa tatttatttta ttatttattt taagttcata ttccatattt attcaagatg  
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 8)

20

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 32
- (2) 3' Untranslated Region - nts 468 - 789

25

Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region  
 5' AUUUUUUUUUUUUUUUUA 3' (SEQ ID NO: 1)

### 6.3. Interleukin 2 ("IL-2")

30

GenBank Accession # U25676:

1 atcactctttaatcacta ctcacattaa cctcaactcc tgccacaatg tacaggatgc  
 61 aactcctgtc ttgcattgca ctaattcttg cacttgtcac aaacagtgcac cctacttcaa  
 121 gttcgacaaa gaaaacaaag aaaacacaggc tacaactgga gcatttactg ctggatttac  
 181 agatgtttt gaatggaaatt aataattaca agaatccaa actcaccagg atgctcata  
 241 ttaagttta catgccaag aaggccacag aactgaaaca gcttcaggtt ctagaagaag  
 301 aactcaaacc tctggagggaa gtgtcaattt tagctcaaag caaaaacttt cacttaagac

361 ccagggactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaaacaa  
 421 cattcatgtg tgaatatgca gatgagacag caaccattgt agaattctg aacagatgga  
 481 ttaccttttg tcaaagcatc atctcaacac taacttgata attaagtgct tcccacttaa  
 5 541 aacatatcag gccttctatt tatttttta aatatttaaa ttttataattt atttgtaat  
 601 gtaggttgc tacctattgt aactattatt ctaaatctta aaactataaa tatggatctt  
 661 ttatgattct tttttaagc cctagggct ctaaaatggt ttaccttatt tatcccaaaa  
 721 atatttatta ttatgtgaa tggtaatagt agtacatgt tagattgggt agtaaaaacta  
 781 ttaataaaat ttgataaata taaaaaaaaaaa aaacaaaaaaaaaa aaaaa (SEQ ID NO: 9)

10 General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 47
- (2) 3' Untranslated Region - nts 519- 825

15 Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region  
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

**6.4. Interleukin 6 ("IL-6")**

20 GenBank Accession # NM\_000600:

1 ttctgccttc gagcccacccg ggaacgaaag agaagctcta tctgcctcc aggagccag  
 61 ctatgaactc cttctccaca aggcgcctcg gtccagttgc cttctccctg gggctgctcc  
 121 tgggttgtcc tgctgccttc cctgccccag taccccccagg agaagattcc aaagatgttag  
 181 ccgccccaaca cagacagccca ctcacccctt cagaacgaaat tgacaaacaa attcggtaca  
 241 tcctgcacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa  
 301 gcagccaaaga ggcactggca gaaaacaacc tgaacccccc aaagatggct gaaaaagatg  
 361 gatgcttcca atctggattc aatgaggaga ctgcctgggt gaaaatcatc actggcttt  
 421 tggagtttga ggtatcaccta ggtacccctc agaacagatt tgagagtagt gaggaacaag  
 481 ccagagctgt gcagatgatg acaaaagtcc tgcacccatgtt cctgcacaaa aaggcaaaaga  
 541 atcttagatgc aataaccacc cctgacccaa ccacaaatgc cagccctgc acgaagctgc  
 601 aggcacagaa ccagtggctg caggacatga caactcatc cattctgcgc agcttaagg  
 661 agttccctca gtccagccctg agggctctc ggccaaatgta gcatggcac ctcagattgt  
 721 tgggtttat gggcattccct tcttcggc agaaacctgt ccactggca cagaacttat  
 781 gttgttctct atggagaact aaaagtatga ggtttaggac actatttaa ttattttt  
 841 ttatataata tttaaaatgt tgaagctgag ttatgtatg taagtcatat ttatatttt  
 901 aagaagtacc acttggaaaca ttatgtat tagtttgaa ataataatgg aaagtggcta

961 tgcagttga atatccttg tttcagagcc agatcattc ttggaaagtg taggcttacc  
 1021 tcaaataaat ggctaactta tacatattt taaagaataa tttatattgt atttatataa  
 1081 tgtataaatg gttttatac caataaatgg catttaaaa aattc (SEQ ID NO: 11)

5

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 62
- (2) 3' Untranslated Region - nts 699 - 1125

10

Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region  
 5' NAUUUAUUUAUUAN 3' (SEQ ID NO: 10)

#### **6.5. Vascular Endothelial Growth Factor ("VEGF")**

15

GenBank Accession # AF022375:

1 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcggcgt  
 61 gcgcgcgcg aaagcgcacag gggcaaagtg agtgcacgtc ttttgggggt gaccggcga  
 121 gcgcggcgtg agccctcccc ctgggatcc cgcagctgac cagtcgcgt gacggacaga  
 181 cagacagaca cgcgcgcgc cccagttac cacctctcc cggccggcg gggacagt  
 241 gacgcggcgg cgagccgcgg gcagggccg gagccgcgc cggaggcgg ggtggaggg  
 301 gtcggagctc gcggcgctgc actgaaactt ttgcgtccaaac ttctgggctg ttctcgcttc  
 361 ggaggagccg tggccgcgc ggggaagcc gagccgagcg gagccgcgag aagtgcgt  
 421 tcggccggg aggagccgca gccggaggag ggggaggagg aagaagagaa ggaagaggag  
 481 agggggccgc agtggcgact cggcgctcg aagccgggct catggacggg tgaggccgc  
 541 gtgtgcgcag acagtgcctc agcgcgcgcg ctccccagcc ctggccggc ctggggccgg  
 601 gaggaagagt agctcgccga ggcgcgcagg agagcgggcc gcggcacagc cggagccgga  
 661 gagggacgcg agccgcgcgc cccggcggg cctccgaaac catgaactt ctgcgtctt  
 721 ggggtgcattt gggccatggcc ttgcgtctt acctccatcca tgccaagtgg tccaggctg  
 781 cacccatggc agaaggagga gggcagaatc atcacgaagt ggtgaagttc atggatgtct  
 841 atcagcgcag ctactgccat ccaatcgaga ccctgggta catctccag ggttaccctg  
 901 atgagatcgatcgatcgatcc aagccatccatgtgtgcgcgtgc gggggctgct  
 961 ccaatgcgaa gggcctggag tgggtgcgcctt ctgaggagtc caacatcacc atgcagat  
 1021 tgcggatcaa acctcaccaa ggccagcaca taggagat gagcttcata cagcacaaca  
 1081 aatgtgaatg cagacccaaag aaagatagag caagacaaga aaatccctgt gggcctgt  
 1141 cagagcggag aaagcattt tttgtacaag atccgcagac gtgtaaatgt tcctgcaaaa  
 1201 acacacactc gcgttgcaag gcggaggcgc ttgagttaaa cgaacgtact tgcagatgt

1261 acaagccgag gcggtgagcc gggcaggagg aaggagcctc ctcagggtt tcgggaacca  
 1321 gatctcttc cagaaagac tgatacagaa cgatcgatac agaaaccacg ctgccgccac  
 1381 cacaccatca ccatcgacag aacagtccct aatccagaaa cctgaaatga aggaagagga  
 5 1441 gactctgcgc agagcacttt gggccggag ggcgagactc cggcggaaagc attccgggc  
 1501 gggtgaccca gcacggccc tcttggaaatt ggattcgcca tttttttttt cttgtgtct  
 1561 aatcaccgag cccggaagat tagagagttt tatttctggg attcctgttag acacaccac  
 1621 ccacatacat acatttatat atatatatat tatatatata taaaaataaa tatcttatt  
 1681 ttatatatat aaaatata tattttttt ttaaataac agtgcataatg ttattgggt  
 10 1741 ctcaactgga tgtatgtac tgctgtggac ttgagttggg agggaaatgt tcccactcag  
 1801 atcctgacag ggaagaggag gagatgagag actctggcat gatctttttt ttgtcccact  
 1861 tggggggcc agggcctct cccctgccc agaatgtgca aggcgggc atggggcaaa  
 1921 atatgaccca gttttggaa caccgacaaa cccagccctg gcgctgagcc tctctacccc  
 1981 aggtcagacg gacagaaaga caaatcacag gttccgggat gaggacacccg gctctgacca  
 15 2041 ggagtttggg gagcttcagg acattgtgt gctttgggg tccctccac atgctgcacg  
 2101 cgcatctcgc ccccaggggc actgcctgga agattcagga gcctggcgg cttcgttta  
 2161 ctctcacctg ctctgagtt gcccaggagg ccactggcag atgtccggc gaagagaaga  
 2221 gacacattgt tggagaagc agcccatgac agcgccctt cctggactc gcccctatcc  
 2281 tcttcctgt ccccttcctg gggcagcc taaaaggacc tatgtctca caccattgaa  
 2341 accactagt ctgtcccccc agaaacactg gttgtgtgt tttttttttt tgacattct  
 20 2401 ccatccccctg gtccctccct tccttcctg aggcacagag agacaggca ggatccacgt  
 2461 gcccattgtg gaggcagaga aaagagaaag tttttatata acggtaactt ttaatatcc  
 2521 ttttttaatt agaaattaga acagtttaattttaaaaaga gtgggtttttt ttttcgttat  
 2581 tcttggtaa tatttaattt caactattt tgagatgtat tttttgtct ctctgtct  
 2641 ttatttgta cgggttttg tatataaaat tcatgtttcc aatctcttc tccctgatcg  
 2701 gtgacagtca ctatgttttcc tggaaacatg atttaattttt gtaacactc agctctgccc  
 2761 tccccgatcc cctggctccc cagcacacat tccttggaaa gagggttca atatacatct  
 2821 acatactata tatataattgg gcaacttgc tttgtgtgtat tatataata tatatgtta  
 2881 tgtatataatg tgatcctgaa aaaataaaaca tcgctattct gttttttata tggtaaaacc  
 2941 aaacaagaaaa aatagagaa ttctacatac taaatcttc tccctttta atttaatata  
 30 3001 ttgttatcat ttatttttg gtgctactgt ttatccgtaa taattgtggg gaaaagatata  
 3061 taacatcagc tcttgcctc tagtgcgtt ttgcgttata ttgcgttata catatattt  
 3121 tttaaacaac gacaaagaaaa tacagatata ttctaaaaaaa aaaaaa (SEQ ID NO: 12)

35 General Target Regions:

(1) 5' Untranslated Region - nts 1 - 701

(2) 3' Untranslated Region - nts 1275 - 3166

### Initial Specific Target Motifs:

5 (1) Internal Ribosome Entry Site (IRES) in 5' untranslated region nts 513 -704  
 5'CCGGGCUCAUGGACGGGUGAGGCAGGCGGUGUGCGCAGACAGU  
 GCUCCAGCGCGCGCUCCCCAGCCCUGGCCGGCCUCGGGCCG  
 GGAGGAAGAGGUAGCUCGCCAGGGCGCCGAGGAGAGCAGGGCCGC  
 CCCACAGCCCGAGCCGGAGAGGGACGCGAGCCCGCGCCCCGGU  
 CGGGCCUCCGAAACCAUGAACUUUCUGCUGUCUUGGGUGCAUU  
 GGAGCCUUGCCUUGCUGCUCUACCUCCACCAUG 3' (SEQ ID NO:  
 13)

10 (2) Group III AU-Rich Element (ARE) Cluster in 3' untranslated region  
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

## 6.6. Human Immunodeficiency Virus I (“HIV-1”)

GenBank Accession # NC\_001802:

1 ggtctctctg gtttagaccag atctgaggcct gggagctctc tggcttaacta gggaaacccac  
61 tgcttaagcc tcaataaaagc ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt  
121 gtgactctgg taactagaga tccctcagac cccttttagtc agtgtggaaa atctcttagca  
181 gtggcgcccc aacaggggacc tgaaagcgaa agggaaacca gaggagctct ctcgacgcag  
241 gactcggctt gctgaagcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc  
301 aaaaatttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa  
361 gcgggggaga attagatcga tgggaaaaaaaa ttcggttaag gccaggggaa aagaaaaaaaaat  
421 ataaataaaa acatatacgta tgggcaagca gggagctaga acgattcgcga gtaatccctg  
481 gcctgttaga aacatcgaaa ggctgttagac aaatactggg acagctacaa ccattccctc  
541 agacaggatc agaagaactt agatcattat ataaatacgat agcaaccctc tattgtgtgc  
601 atcaaaggat agagataaaaa gacaccaagg aagctttaga caagatagag gaagagcaaa  
661 acaaaaagtaa gaaaaaagca cagcaagcag cagctgacac aggacacacgc aatcaggta  
721 gccaaaattt ccctatagtg cagaacatcc aggggcaat ggtacatcag gccatatcac  
781 ctagaacttt aatgcattgg gtaaaagtag tagaagagaa ggcttcagc ccagaagtga  
841 tacccatgtt ttccagcattt tcagaaggag ccacccacaca agatttaaac accatgtta  
901 acacagtggg gggacatcaa gcagccatgc aatgttaaa agagaccatc aatgaggaaag  
961 ctgcagaatg ggatagatgt catccagtgc atgcagggcc tattgcacca ggccagatgt  
1021 gagaaccaag gggaaagtgc atagcaggaa ctactagtac ctttcaggaa caaataggat  
1081 ggatgacaaa taatccaccc atcccgatgt gagaattta taaaagatgg ataatccctgg

1141 gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac  
1201 caaaggaacc cttagagac tatgtagacc ggttctataa aactctaaga gccgagcaag  
1261 cttcacagga ggtaaaaat tggatgacag aaaccttgtt ggtccaaaat gcgAACCCAG  
5 1321 atttaagac tattttaaa gcattggac cagcggctac actagaagaa atgatgacag  
1381 catgcaggg agtaggagga cccggccata aggcaagagt ttggctgaa gcaatgagcc  
1441 aagtaacaaa tttagtacc ataatgatgc agagaggcaa ttttaggaac caaagaaaga  
1501 ttgttaagtgt ttcaattgtt ggaaagaag ggcacacacgc cagaaattgc agggccctta  
1561 ggaaaaaggg ctgtggaaa tgtggaaagg aaggacacca aatgaaagat tgtactgaga  
10 1621 gacaggctaa ttttttaggg aagatctggc cttcctacaa gggaggcca gggattttc  
1681 ttcagagcg accagagcca acagccccac cagaagagag cttaggtct gggtagaga  
1741 caacaactcc ccctcagaag caggagccga tagacaagga actgtatcct ttaacitccc  
1801 tcaggtcact cttggcaac gaccctcgta cacaataaag atagggggc aactaaagga  
1861 agctcttata gatacaggag cagatgatac agtattagaa gaaatgagg tgcaggaaag  
15 1921 atgaaacca aaaatgatag gggaaattgg aggtttatc aaagtaagac agtattatca  
1981 gatactcata gaaatctgtg gacataaagc tataggtaca gtatttagtag gacctacacc  
2041 tgtcaacata attgaaagaa atctgttgc tcagattgtt tgcactttaa atttcccat  
2101 tagccattt gagactgtac cagtaaaattt aaagccagga atggatggcc caaaagttaa  
2161 acaatggcca ttgacagaag aaaaataaaa agcattagta gaaattgtt cagagatgga  
2221 aaaggaaggg aaaatttcaa aaattggcc tgaaaatcca tacaatactc cagttttgc  
2281 cataaagaaa aaagacagta ctaaatggag aaaaattgtt gatttcagag aacttaataa  
2341 gagaactcaa gacttctggg aagttcaattt aggaatacca catcccgag ggtaaaaaaa  
2401 gaaaaatca gtaacagtac tggatgtggg ttagtgcata ttttcagttt ccttagatga  
2461 agacttcagg aagtatactg catttaccat acctgtata aacaatgaga caccaggat  
25 2521 tagatatcag tacaatgtgc tccacaggatggaaagga tcaccagcaa tattccaaag  
2581 tagcatgaca aaaatcttag agcccttttag aaaaacaaaat ccagacatag ttatctatca  
2641 atacatggat gattgtatg taggatctga ctttagaaata gggcagcata gaacaaaaat  
2701 agaggagctg agacaacatc tggatgtgg gggacttacc acaccagaca aaaaacatca  
2761 gaaagaacct ccattccctt ggtgggtta tgaactccat cctgataaat ggacagtaca  
2821 gcctatagtg ctggcagaaa aagacagctg gactgtcaat gacatacaga agttatggg  
30 2881 gaaattgaat tggcaagtc agtttaccc agggattaaa gtaaggcaat tatgtaaact  
2941 ccttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctaga  
3001 actggcagaa aacagagaga ttctaaaaga accagtcat gggacttacc acaccatc  
3061 aaaagactta atacagaaatc tacagaagca gggcaaggc caatggacat atcaaattta  
3121 tcaagagcca tttaaaaatc tgaaaacagg aaaatgtca agaatgaggg gtgcacacac  
35 3181 taatgtatc aaacaattaa cagaggcagt gaaaaata accacagaaa gcatagtaat





7441 ggctattgag ggcacacagc atctgttgc actcacagtc tggggcatca agcagctcca  
 7501 ggcaagaatc ctggctgtgg aaagataacct aaaggatcaa cagctccctgg ggatttgggg  
 7561 ttgccttgaa aaactcattt gcaccactgc tgtgccttg aatgcttagtt ggagtaataa  
 5 7621 atctctggaa cagatttggaa atcacacgac ctggatggag tggacagag aaattaacaa  
 7681 ttacacaaggc ttaataact ccttaatttga agaatcgcaaa accagcaag aaaagaatga  
 7741 acaagaatta ttggaattttag ataaatgggc aagtttgtgg aattggtttta acataacaaa  
 7801 ttggctgtgg tataaaaaat tattcataat gatagttagga ggcttggtag gttttaagaat  
 7861 agttttgttgcgtactttcta tagtgaatag agttaggcag ggatattcac cattatcgtt  
 10 7921 tcagacccac ctcccaaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg  
 7981 tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg  
 8041 ggacgatctg cggagccctgt gcctcttcag ctaccaccgc ttgagagact tactcttgat  
 8101 ttaacgagg atttgttggaaac ttctggacg caggggggtgg gaagccctca aatattttgt  
 8161 gaatctcttca cagtatttggaa gtcaggaact aaagaatagt gctgttagct tgctcaatgc  
 15 8221 cacagccata gcagtagctg agggacaga tagggttata gaagttagtac aaggagctg  
 8281 tagagctatt cgccacatac cttagaagaat aagacaggc ttggaaagga ttttgctata  
 8341 agatgggtgg caagtggtca aaaagttagtg tgaatggatg gcctactgttta agggaaagaa  
 8401 tgagacgagc tgagccagca gcagataggg tggagcagc atctcgagac ctggaaaaac  
 8461 atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttgc tggctagaag  
 20 8521 cacaagagga ggaggagggtg ggtttccag tcacacctca ggtacccatagaccaatga  
 8581 cttacaaggc agctgttagat cttagccact tttaaaaga aaaggggggaa ctggaaaggc  
 8641 taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct  
 8701 actccctga ttagcagaac tacacaccag ggccagggtt cagatatcca ctgacccttg  
 8761 gatgggtgcta caagcttagta ccagttgagc cagataagat agaagaggcc aataaaggag  
 25 8821 agaacaccag ctgttacac cctgtgagcc tgcatggat ggatgaccgc gagagagaag  
 8881 tgtagagtg gaggttgac agccgcctag catttcatca cgtggcccg gagctgcac  
 8941 cggagtactt caagaactgc tgacatcgag ctgctacaa gggacttcc gctggggact  
 9001 ttccaggag gcggtggctg ggcggactg gggagtggcg agccctcaga tcctgcata  
 9061 aagcagctgc ttttgcctg tactgggtct ctctggtag accagatctg agcctggag  
 30 9121 ctctctggct aactaggaa cccactgttt aagcctcaat aaagcttgc ttagtgc  
 9181 c (SEQ ID NO: 14)

Initial Specific Target Motifs:

(1) Trans-activation response region/Tat protein binding site - TAR RNA - nts 1

35 - 60

“Minimal” TAR RNA element

5' GGCAGAUCUGAGCCUGGGAGCUCUCUGCC 3' (SEQ ID NO: 15)

(2) Gag/Pol Frameshifting Site - "Minimal" frameshifting element

5'

5 UUUUUUAGGGAAGAACUGGCCUUCCUACAAGGGAAGGCCAGG  
GAAUUUUCUU 3' (SEQ ID NO: 16)

### 6.7. Hepatitis C Virus ("HCV" - Genotypes 1a & 1b)

GenBank Accession # NC\_001433:

10 1 ttggggcga cactccacca tagatcaact ccctgtgagg aactactgtc ttcacgcaga  
61 aagcgcttag ccatggcggtt agtatgagtg ttgtgcagcc tccaggaccc cccctcccg  
121 gagagccata gtggctcgcg gaaccggta gtacaccgga attgccagga cgaccgggtc  
181 ctttcttggaa tcaaccggct caatgcctgg agatttgggc gtgcggccgc gagactgcta  
241 gccgagtagt gtgggtcgca gaaaggcctt gtggtactgc ctgatagggt gcttgcgagt  
301 gccccgggag gtctcgtaga ccgtgcatca tgagcacaaa tcctaaacct caaagaaaaaa  
361 ccaaacgtaa caccacccgc cgcccacagg acgttaagtt cccggcggt ggtcagatcg  
421 ttggtgagggtt acactgttgc cgcgcaggg gccccagggtt ggggtgcgc gcgacttagga  
481 agactccgaa gcggtcgaa cctcgtggaa ggcgacaacc tatcccaag gtcggccggc  
541 ccgagggttag gaccctggctt cagccgggtt acccttggcc cctctatggc aacgagggtt  
601 tggggggc aggatggctc ctgtcacccccc gtggctctcg gcctagttgg gccccacag  
661 accccggcg taggtcgctt aatttggta aggtcatcgat tacccttaca tgcggcttcg  
721 ccgacccat ggggtacatt ccgcctgtcg ggcggccctt agggggcgctt gccaggccc  
781 tggcacatgg tggccgggtt ctggaggacg gcgtaacta tgcaacaggg aatctggcc  
841 gtgtcgcttt ctctatcttc ctcttagctt tgctgtcttg ttgaccatc ccagcttcgg  
901 cttagggtt ggcacatgg tccggatataccatgtc acgtcgactc tccaaactcaa  
961 gtattgtta tgaggcagcg gacatgtca tgcacacccccc cgggtgcgtg ccctgcgtcc  
1021 gggagagtaa ttctcccggt tgctgggttag cgctcactcc cacgctcgcc gccaggaaca  
1081 gcagcatccc caccacgaca atacgacgac acgtcgattt gtcgttggg gcggtgcct  
1141 tctgtccgc tatgtacgtt gggatcttcc gggatcttcc tttctcgcc tcccgatgt  
1201 tcaccttcc acctcgccgg tatgagacgg tacaagattt caatgtca atctatcccg  
1261 gcccacgtatc aggtcaccgc atggcttggg atatgtatgtt gaaactggta cctacaacgg  
1321 cccttaggtt atcgcatca ctccggatcc cacaagccgt cgtggacatg gtggcgcccc  
1381 cccactgggg tggccatcg ggccttgcctt actattccat ggtggggaaac tgggttaagg  
1441 tctgtattgtt gatgtactc tttgtcggtt tgacggca caccacgtg acagggggaa  
1501 gggtagcctc cagcacccag agccctcggtt cctggcttc acaaggccca ttcagaaaa  
1561 tccaaactcggtt gaacaccaac ggcagctggc acatcaacag gaccgtctg aattgcaatg

1621 actccctcca aactgggttc attgctgcgc tggctacgc acacagggtc aacgcgtccg  
 1681 ggtccccaga gcgcattggct agctgcgcgc ccatcgatga gttcgctcag ggggtgggtc  
 1741 ccatcactca tggatatgcct gagagctcgg accagaggcc atattgctgg cactacgcgc  
 5 1801 ctgcaccgtg cgggatcgtg cctgcgtcgc aggtgtgtgg tccagtgat tgcttcactc  
 1861 cgagccctgt tggatgtgggg acgaccgatc gttcggcgc tcctacgtat agctgggggg  
 1921 agaatgagac agacgtgctg ctacttagca acacgcggcc gcctcaaggc aactggtttgc  
 1981 ggtcacgtg gatgaacagc actgggttca ccaagacgtg cggggccct ccgtgcaaca  
 2041 tcgggggggt cggcaacaac accttggctc gccccacggta tgcgtccgg aagcaccccg  
 10 2101 aggccactta cacaaggatgt ggctcgggc cctgggtgac acccagggtgc atgggtgact  
 2161 acccatacag gctctggcac taccctgcata ctgttaacattt taccgtctt aaggtcagga  
 2221 tggatgtggg gggcggtggag cacaggctca atgctgcgtg caattggact cgaggagagc  
 2281 gctgtgactt ggaggacagg gataggtcgtg aactcagccc gctgctgtg tctacaacag  
 2341 agtggcagat actgcccgtt tccttcacca ccctaccggc cctgtccact ggcttgatcc  
 15 2401 atcttcacccg gaacatcggtg gacgtcaat acctgtacgg tatagggtcg gcagttgtct  
 2461 ccttgcaat caaatggag tatatcctgt tgctttccct tcttctggcg gacgcgcgcgc  
 2521 tctgtgcctg ctgtggatg atgctgcgtg tagcccgaggc tgaggccacc ttagagaacc  
 2581 tgggtggctt caatgcggcg tctgtggcccg gagcgcatgg ccttctctcc ttctcggt  
 2641 tcttcgcgc cgcctgggtac atcaaaggca ggctggtccc tggggcggca tatgctctct  
 2701 atggcgtatg gcccgtcctc ctgtcttcg tggcttacc accacgagct tatgccatgg  
 2761 accgagagat ggctgcgtc tgccggaggcg cggttttgtt aggtctggta ctcttgacct  
 2821 tgcaccata ctataagggtg ttccctcgcta ggctcatatg gtggttacaa tattttatca  
 2881 ccagagccga ggccgcacttg caagtgtggg tccccctct caatgtcgg ggaggccgcgc  
 2941 atgcacatcat cctccctaca tgcgcggtcc atccagagct aatcttgac atcaccaaac  
 3001 tcctgtcgc catactcggt ccgcctatgg tgctccaggc tggcataact agagtgcgcgt  
 3061 actttgtacg cgctcagggg ctatccgtg catgcgtt agtgcggaaag gtcgctggag  
 3121 gccactatgt ccaaattggcc ttcatgaagc tggccgcgcgt gacaggtacg tacgtatatg  
 3181 accatcttac tccactgcgg gattggccc acgcgggcct acgagacattt gcgggtggcag  
 3241 tagagcccggt cgtttctct gacatggaga ctaaaactcat cacctggggg gcagacaccg  
 3301 cggcgtgtgg ggacatcatc tcgggtctac cagtcctccgc ccgaaggggg aaggagatac  
 3361 ttcttaggacc ggccgatagt ttggagagc aggggtggcg gctcctgcgc cctatcacgg  
 3421 cctattccca acaaacgcgg ggcctgcgtg gctgtatcat cactgcctc acaggtcggg  
 3481 acaagaacca ggtcgatggg gaggttcagg tgctctccac cgcaacgcaa tcttcctgg  
 3541 cgacccgtcgtt caatggcgatg tggatgtggaccg tctaccatgg tggccgcgt aagaccctgg  
 3601 ccggcccgaa ggtccaaatc accaaatgtt acaccaatgtt agaccaggac ctcgtcggct  
 35 3661 ggccggcgcc ccccgggcgcc cgctccatga caccgtcgcac ctgcggcagc tcggacctt



5821 gcgcggcat tgccggtgcg gccgtggca gcataggctc cgggaaggta cttgtggaca  
5881 ttctggggg ctatggggcg ggggtggctg ggcactcg ggccttaag gtcatgagcg  
5941 gcgagatgcc ctccactgag gatctggta atttactccc tgccatcctt ttcctggcg  
5 6001 ccctggttgt cggggcgtg tgccgagcaa tactgcgtcg gcacgtggc cggggagagg  
6061 gggctgtgca gtggatgaac cggctgatag cgttcgttc gggggtaac cacgtctccc  
6121 ccacgcacta tggcccgag agcgacgccc cggcgctgt tactcagatc ctctccagcc  
6181 ttaccatcac tcagttgctg aagaggcttc atcagtggat taatgaggac tgctccacgc  
6241 ctgttccgg ctgcgtgcta aaggatgtt gggactggat atgcacggtg ttgagtgact  
10 6301 tcaagacttg gctccagtcc aagctccgtc cgcggttacc gggactccct ttccgtcat  
6361 gccaacgcgg gtacaaggga gtctggggg gggatggcat catgcaaacc acctgcccatt  
6421 gtggagcaca gatcaccggaa catgtcaaaa atggctccat gaggattttt gggccaaaaaa  
6481 cctgcagcaa cacgtggcat ggaacattcc ccatcaacgc atacaccacg ggccctgca  
6541 cgcctcccc agcgccgaac tattccaggg cgctgtggcg ggtggctgct gaggagtacg  
15 6601 tggaggttac gccccggggg gatttccact acgtgacggg catgaccact gacaacgtga  
6661 aatgcccattt ccagggttcca gcccctgaat ttacgttccat ggtggatgga gtacgggttc  
6721 acaggtatgc tccagtgtgc aaacctctcc tacgagagga ggtcgatcc caggtcgggc  
6781 tcaaccatg cctgggtggg tcacagtc catgtgagcc cgaaccggat gtggcagtgc  
6841 tcactccat gtcaccgac ccctctcata ttacagcaga gacggccaag cgtaggctgg  
6901 ccaggggttc tccccctcc ttggccagct cttagtgc ccagttgtct ggccttctt  
20 6961 tgaaggcgac atgtactacc catcatgact cccggacgc tgacccatc gaggccaacc  
7021 tcctgtggcg gcaggagat ggcgggaaca tcacccgtgt ggagtcagaa aataaggtgg  
7081 taatcccttgcg ctcttcgtat ccgttccggg cgggtggagga tgagaggaa atatccgtcc  
7141 cggcgagat cctgcgaaaaa cccaggaat tccccccatc gttggccata tggcacgccc  
25 7201 cggattacaa ccctccactg cttagtgcctt ggaaggaccg ggactacgtc ccccccgtgg  
7261 tacacgggtg cccttgcgc tctaccaagg ccccccaat accaccatcc cggaggaaga  
7321 ggacgggtgt cctgacagag tccaccgtgt ctctgcctt ggcggagctc gctactaaga  
7381 cccttggcag ctccgggtcg tcggccgttg acagcggcac ggcgactggc cctccgtatc  
7441 aggccctccga cgacggcgac aaaggatccg acgttggatc gtactccatc atgccccccc  
7501 tcgagggaga gccaggggac cccgaccata ggcgggttc ttggcttacc gtgaggggg  
30 7561 aagctggta ggacgtgc tgctgtcaa tgcctatac atggacagggt gccttgatca  
7621 cgcctatgcgc tgccggaggag agcaagttgc ccatcaatcc gttgagcaac tcttgcgtc  
7681 gtcaccacag tatggctac tccacaacat ctgcagcgc aagtctggc cagaagaagg  
7741 tcacccatgaa cagactgcaat gtcctggacg accactaccg ggacgtgctc aaggagatga  
7801 aggcaaggc gtccacagt aaggcttaggc ttcttatctat agaggaggcc tgcaaactga  
35 7861 cgcctccaca ttccggccaaa tccaaatttg gtcacggggc gaaggacgtc cggagccat

7921 ccagcaggc cgtcaaccac atccgctccg tgtgggagga cttgctggaa gacactgaaa  
 7981 caccaattga taccaccatc atggcaaaaa atgagggttt ctgcgtccaa ccagagaaag  
 8041 gaggccgcaa gccagctcgc cttatcgat tcccgacacct gggggtatgt gtatgcgaga  
 5 8101 agatggccct ttacgacgtg gtcctccaccc ttccctcaggc cgtgatggc ccctcatacg  
 8161 gattccaga ctctcctggg cagcgggtcg agttcctgtt gaatacctgg aatcaaaga  
 8221 aatgccctat gggcttctca tatgacaccc gctgcttga ctcaacggtc actgagaatg  
 8281 acatccgtac tgaggaatca atttaccaat gtgtgactt ggcccccggaa gccaggcagg  
 8341 ccataaggc gctcacagag cggctttatg tcgggggtcc cctgactaat tcgaaggggc  
 10 8401 agaactgcgg ttatgcggg tgccgcgcaa gtggcgtgct gacgactagc tgcggcaaca  
 8461 ccctcacatg ttacttgaag gccactgcgg cctgtcgagc tgcaaagctc caggactgca  
 8521 cgtatgcgtgta acggagac gacccgtcg ttatctgtga gagtgcgggaa acccaggagg  
 8581 atgcggcggc cttacgagcc ttacggagg ctatgactag gtattccgcc ccccccgggg  
 8641 acccgccccca accagaatac gacttggagc tgataacgtc atgctccctc aatgtgtcg  
 15 8701 tcgcgcacga tgcatccggc aaaagggtgt actacctcac ccgtgacccc accacccccc  
 8761 tcgcacggc tgcgtggag acagtttagac acactccagt caactcctgg ctaggcaata  
 8821 tcatcatgtt tgcggccacc ctatggcga ggtatgttctt gatgactcat ttcttccta  
 8881 tccttcttagc tcaggagcaaa ctgaaaaag ccctggattt tcagatctac ggggccttt  
 8941 actccattga gccacttgac ctacccaga tcatgttacg actccatgtt ctatgcgtat  
 20 9001 ttcaacttca cagttacttccatgttacg tcaatagggt ggcttcatgc ctcaggaaac  
 9061 ttgggttacc gccttgcga gtctggagac atcggccag aagtgtccgc gctaagctac  
 9121 tgtcccgagg ggggagggtt gcccacttgcg gcaagtaccc ttcaacttgg gcaatggaa  
 9181 ccaagcttaa actcaacttca atccggctt cgtcccgatc agacttgcgg ggcttgcgtt  
 9241 ttgtgttta caacgggggaa gacatatac acagcctgtc tgcgtccggc ccccggttgg  
 9301 tcatgttgc ctacttcata ctttctgtt gggtagggcat tcatgttgc cccaaaccgg  
 25 9361 gaacggggag ctaaccactc caggccataa ggccattccc tttttttttt ttc (SEQ ID NO: 17)

General Target Region:

5' Untranslated Region - nts 1 - 328 - Internal Ribosome Entry Site (IRES):

30 5' UGGGGGGCGACACUCCACAUAGAUCAUCACUCCCCUGUGAGGAACUACUGUCU  
 UCACGCAGAAAGCGUCUAGCCAUGGCGUUAGUAUGAGUGUUGUGCAGCCUC  
 CAGGACCCCCCUCCGGAGAGCCAUAUGUGGUCUGCGGAACCGGUGAGUAC  
 ACCGGAAUUGCCAGGACGACCGGGGUCCUUUCUUGGAUCAACCCGCUAAUGC  
 CUGGAGAUUUGGGCGUGCCCCCGCGAGACUGCUAGCCGAGUAGUGUUGGGU  
 35 CGCGAAAGGCCUUGUGGUACUGCCUGAUAGGGUGCUUGCAGUGCCCCGGG  
 AGGUCUCGUAGACCGUGCAU3' (SEQ ID NO: 18)

### Initial Specific Target Motifs:

5 (1) Subdomain IIIc within HCV IRES - nts 213 - 226  
5'AUUUGGGCGUGCCC3' (SEQ ID NO: 19)  
(2) Subdomain IIId within HCV IRES - nts 241-267  
5'GCCGAGUAGUGUUGGGUCGCGAAAGGC3' (SEQ ID NO: 20)

## 6.8. Ribonuclease P RNA (“RNaseP”)

### GenBank Accession #s

## X15624 Homo sapiens RNaseP H1 RNA:

1 atggcgagg ggaagctcat cagtggggcc acgagctgag tgcgtccctg cactccact  
61 ccatgtccct tgggaagggtc tgagactagg gccagaggcg gccctaacag ggctctccct  
121 gagcttcagg gaggtgagtt cccagagaac ggggctccgc gcgaggcag actgggcagg  
181 agatgccgtg gaccccgccc ttggggagg ggcccgccgg atgcctccct tgccggagct  
241 tggAACAGAC tcacggccag cgaagtgagt tcaatggctg aggtgaggtt ccccgccagg  
301 gacctataa cccaaattcag accactctcc tccggccatt (SEQ ID NO: 21)

## U64885 *Staphylococcus aureus* RNaseP (rrnB) RNA:

20 1 gagggaaagtc cgggctcaca cagtctgaga tgattgtagtttcgtgcgtt gatgaaacaa  
61 taaaatcaagg cattaatttg acggcaatga aatatcctaa gtcttcgtatggatagag  
121 taatttggaaa gtgccacagt gacgttagctt ttatagaaat ataaaaggtg gaacgcggta  
181 aaccctcgatgtgagcaatc ccaaatttggtaggagcactt gttaacggatattcaacgtt  
241 taaaacgagac acacttcgcgaaaatgaaatgtgttagacagatggttatca cctgagtacc  
301 agtgtgactatgtgcacgtga tgagtacgtatggacacaaacgcggcttat (SEO ID NO: 22)

M17569 Escherichia coli RNA component (M1 RNA) of ribonuclease P (rnpB) gene:

30 1 gaagctgacc agacagtgcg cgcttcgtcg tcgtcctt cgggggagac gggcggaggg  
61 gaggaaagtc cgggctccat agggcagggt gccaggtaaac gcctgggggg gaaacccacg  
121 accagtgcaa cagagageaa accgcccgtg gcccgcgcaa gcgggatcag gtaagggtga  
181 aagggtgcgg taagagcgcgca cgcgcggct ggtaacagtc cgtggcacgg taaactccac  
241 ccggagcaag gccaaatagg ggttcataag gtacggcccg tactgaaccc gggtaggctg  
301 cttgagccag tgagcgattg ctggcctaga tgaatgactg tccacgacag aaccggctt  
361 atcggtcagt ttaccc (SEQ ID NO: 23)

Z70692      *Mycobacterium tuberculosis* RNaseP (rnpB) RNA:

1 ccaccggta ccatggcc gaccatggcc ccacaatagg gcccgggaga cccggcgta  
 61 gtggggcg gcacggtcag taacgtctgc gcaacacggg gttgactgac gggcaatac  
 121 ggctccatag cgctggccgc ggatacagta aaggagcatt ctgtgacgga aaagacgccc  
 181 gacgacgtc tcaaacttgc caaggacgag aaggtcaat atgtcgacgt ccggttctgt  
 241 gacctgcctg gcatcatgca gcacttcacg attccggctt cggccttga caagagcgtg  
 301 ttgacgacg gcttggccctt tgacggctcg tgcattcg ggttccagtc gatccacgaa  
 361 tccgacatgt tgcatttcc cgtatcccgag acggcgcgca tcgacccgtt cccgcggcc  
 421 aagacgctga atatcaactt ctttgtcac gaccgttca ccctggagcc gtactcccgc  
 481 gaccgcgcga acatgcggc caaggccgag aactacactga tcagcactgg catgcggac  
 541 accgcatact tcggcgccga ggccgagttc tacatttcg attcggttag ctgcactcg  
 601 cgcgcacacg gctcttcata cgaggtggac gccatctcg ggtggtagaa caccggcg  
 661 ggcacccgagg ccgcacggcag tcccaaccgg ggctacaagg tccgcacaa gggcggtat  
 721 ttcccaactgg ccccaacga ccaatacgtc gacgtcgccg acaagatgt gaccaacctg  
 781 atcaactccg gcttcatctt ggagaaggc caccacgagg tggcagcgg cggacaggcc  
 841 gagatcaact accagttcaa ttgcgtctg cactggcccg acgacatgca gtttacaag  
 901 tacatcatca agaacaccgc ctggcagaac ggcaaaacgg tcacgttcat gccaagccg  
 961 ctgttcggcg acaacgggtc cggcatgcac tgtcatcagt cgctgtggaa ggacggggcc  
 1021 ccgcgtatgt acgacgagac gggttatgcc ggtctgtcg acacggcccg tcattacatc  
 1081 ggccggctgt tacaccacgc gccgtcgctg ctggccttca ccaacccgac ggtgaactcc  
 1141 tacaagcggc tggttcccg ttacgaggcc ccgatcaacc tggctatag ccagcgcaac  
 1201 cggcggcat gctgtcgcat cccgatcacc ggcagcaacc cgaaggccaa gcccgtggag  
 1261 ttccgaagcc ccgactcgac gggcaaccgg tatctggcg tctcgccat gctgtatggca  
 1321 ggcctggacg gatcaagaa caagatcgac ccgcaggcgc cgctgacaa ggatcttac  
 1381 gagctgcgcg cggaaaggagc cgcgagttac ccgcagactc cgaccagct gtcagatgt  
 1441 atcgaccgtc tcgaggccga ccacgaatac ctcaccgaag gagggtgtt cacaacgac  
 1501 ctgtcgaga cgtggatcg ttcaagcgc gaaaacgaga tcgagccgtt caacatccgg  
 1561 ccgcattccct acgaattcgc gctgtactac gacgtttaag gactttcgc agtccgggtg  
 1621 tagagggagc ggcgtgtcg tggcaggccg ggcgtcgagg ttttcgtt ggtgacgggt  
 1681 gcccggcaacg gcccggcgcac caccgctcg gaaagccgtt ttaagaacgt tcaaggacgt  
 1741 ttccggccggg tgccacaacc cgttggccaa tcatctcccg accgcccggc ggggtgtt  
 1801 tcacatgcgc cggaaactcaa gccacgtcg cgcggccggc tgcgtcgcc gcccgttgc  
 1861 gttaaatgtc gggattcgt cgtggccggc ggcgtccacg ctgaccaacg gggcagtc  
 1921 ctcccaaca ctttgcgcac taccgcctt gcccggccg tcacccgtt gtagttgtcc  
 1981 aggaattcccc caccgtcgac gtttgcgcag cggccggcga cccgcggccg attgagctgg

2041 cggccgggtc cggcagctg gtcggggc ttgccgcga ccaacaccag cgcggtgcgg  
 2101 gcccgggtgg cggtcagcca ggcctgacgg agcagctcca cgtcgctgc ggaaaccaga  
 2161 tcggcggccg cgatgacatc cagggattgc agcgtcgagg ttttgtcag ggcgggaacc  
 5 2221 tggtcgcgtat gctgttagctg cagcaactgc acggccattt cgtatgcggc cagtccggcc  
 2281 cggcccagtt tgggtgtgt gtggggcgc gcacccgcgc gcaaccgctc ggactcgata  
 2341 cgggccttga tgcggcgaat ctcgcgcacc gactcagcgg acacaccgtc gggccgat  
 2401 cgcgttttgt cgaccatccg taggaatcgc tgacccaaact cggcatcgcc ggcaaccgc  
 2461 tgtgcgcgtat gcagggcctg gatctccat ggctgtgcgc actgctcgta gtatgcggcg  
 10 2521 taggacccca ggggtcggac cagcggaccg ttgcggccct cgggtcgcaa attggcgtcg  
 2581 agctccagcg gcggtatcgac gctgggtgtc cccagcagcg cccgaaccgc ctggcgatc  
 2641 gatgtcgacc atttcaccgc cctgtcatcg tcgacgcgg tggccggctc acagacgaac  
 2701 atcacgatgg catccgaccc gtagccaaac tcggcaccac ccagccgacc catggcgatg  
 2761 acccgcatgg ccgcggggc gcatcgatcg tcggaaaggc tggccggat catgacgtcc  
 15 2821 agcgcggccct gcagcaccgc caccacacc gacgtcaacg cccggcacac ctgggtgacc  
 2881 tcgagcaggc cgagcaggc cgccgaaccg atgcgggcca gctctcgacg acgcagcg  
 2941 cgcgcggcccg cgatggcccg ctccgggtcg gggtagcggc tcgcccggc gatcagcgcc  
 3001 cgagccacgg cggcgggcgc ggtctcgacg agctcgggc ccgcaggccc gtccctgtac  
 3061 tgctggatga cccgcggcgc gcatcgatcaac agatccggca catacgccga ggtacccaa  
 3121 acatcgatga gcccgtggc caccgcggc ttgtcccgca gctggccag gtaccagctt  
 20 3181 tcggtgccca ggcctcact gagccgcgg taggcagca gtccgcgtc gggatcgaaaa  
 3241 gcatacgaca tccagtccag cagccgtggc agcagcaccg actgcaccccg tccgcggcc  
 3301 ccgcgttgc tgaccaacgc cgacatgtgt tcaacgcgg tctgcggcctc ctctgtac  
 3361 agcgcggcca gcccggccccc cgcggccctcc aacgtcatgc cgtggcgat ctccaaacc  
 3421 gtcggccga tcgattccag cagcggttga tagaagagg tgggtgttaa ctgcacacc  
 25 3481 cgcacgttct gcttcttgcg ttccctccgc agcaccggc cgcacatcg tccgcgcatt  
 3541 ggccggatgt gggccgcgc cgcgcggccag cgcactgcct cctcgcttc gggatcgaaa  
 3601 agcaggtggg tgcgttgcg cgcgtcaac tgcagtcgg gctcgagcag cctgaggaac  
 3661 tcatacgacg cggcatgtt cgcgcgtcc tcacgcccga ttagccgc ttgcggcaac  
 3721 gcccggccatg cgccaccgt ggacgccacc cgtaacgact cgtcgatcg ggcacatgaa  
 3781 agctcgatg gctgtacggc gaaactccacg tcgcgcataatc cgcgcgtgc gaggttgagc  
 3841 tcggccgc ggcacatccgc gggcaccgc tgctccaccc ggcgcgcgc ggcctgcacc  
 3901 tcgaccacaa agcttcgcg ctcgcaggct cgcacacca tcggcatcaa ggcggtcagg  
 3961 taacgtcgatgc caagtccgc gtcgcgcataatc actggccgtt cttcagccaa cgcctgaa  
 4021 tcccggtct tggccagcg ctggtagtag gcatgtgcg actcgatcgatcg accggacc  
 35 4081 tcccggttgc gcccgtccgg acgcaggcgc gcatccaccc cggaaaaaggc cgcggaggcc

4141 acccgcatca tctcgctggc cacgcgcgcg ttgcgcgggt cggagcgctc ggcaacgaat  
4201 atgacatcg a c t c g c t g a c g t a g t t c a g t t c g c g c a c c g a c t t g c c a t c g c g a t g t  
4261 accgccaggc gcgg tggcgg gtgc tgc cccg cacacgctg c t c g g c c a c g c a g c g c c  
4321 g c g c c a g a g c g g c t c c g c g c g c t c c g c a g g c t g c g g c c a c c a c g g t g a a t g g c a g c  
4381 accgg t c g t c t c g a c c g t c g c g c a g g t c a g g c t g c g g c c a c g t a t t g a c t g c g  
4441 c g g t a c t g g g t t c g c a a t c g g t c a c g a c g a g c g a g c c c g g c a t a c c t c c g a t t c c t c g a c g  
4501 c a c t c g a c g a c g a c g c t g c a g c t g g t c a t g g a c g g c a g t g a c c t t g c a c g  
4561 a a t t c c a g g a c t g c g g a t g g c a c c a g g t c a g c c a g c g a c g a g c c c a g c  
4621 a c c g a g a a c a c g c c c c g c g a c a g a c t g c g t c g c g c a g c a g c a g c c g c g t t g a c t c g t c c  
4681 c a t c c g g t g t c t g a t t c t c g a c a g c g c g g a t c a a g g c g c g a c g c g c g g c a t c g g c g t c c  
4741 g g a g c g c g t g a c a g c g a c c a g c a g g t c g c g c c t g a t c c t c g t g c c g a t c c c a c  
4801 c c c a g c t g a g c c a g c g c t c a c c a g c a g g g g g t c a a c t a t c c g a g c c g g c c a a c g c t g  
4861 g g c a a c t t c g g c c g t g c g g c a g g t t t g g t c a c g a c c a g c g g t a g c g a a a g c g c g  
4921 t c g g c g t c g g a t c a a c c g g t a g a t c t g g g c a c g a c a g t g c g c a g c t g t a t  
4981 g g c g t g a c g t g g t c g g t a a t c c g c c a c t c g c g c t g t g c g c a a g a a g t c a  
5041 a a a a c g t g c t c c c c a a g g c c t c c g c g a c g a t t c g g a g g c c t c a t g g c g c a g c g c a  
5101 c t a t c c a a a c t g g a c g g c a a t c c g g f a c c c a t c g c t c g c t c g g t g a g g  
5161 t c c c a t a c g t g t c g g c c t g c g g g c c c a g c a c g t a a c c t t c t a c a c c c g c a a t  
5221 c c c g g g c c a g c a c g g c g a a t g t c a g a t g g a t t g c a c g c c a a t c a g g c t g c g t  
5281 a c t t c g a c c c g c g a c g a g t t g t g c g c g t g t a c a t g g a c c c g c a c t a g g g c g  
5341 g a t c g g t t g g c g g c c c a g c a c g g c c g t g g g c g c t c g c c c c t g c a c c a g c c g c  
5401 t t g t a a g a g t g a c c c a c t g a t t g a c c g c g t g a t c a c g c t g c t c a g g a t t g  
5461 c c g g c g a t g a a c g a t t a c c a t t c c g a c a g t g c a g c g a t c a t c a g c g t g g a a c  
5521 g c g t g a c a t c a c c c t c g a a c a g g c t a t g t g g a t t g c a t g c c g a g c c g g t g c t g  
5581 c c g a a t g g c t t g g c a t g a a c g a c g c c c g g c g c a c t t t g a t g  
5641 a c g t a g c g g a a g g t c a t c a c g a t t g t c a g c a g a g a c g t g g c a a a c c g a g g t c g  
5701 a t c c t c g c t g g c c g g g t g c g c t c g a a t t g g a c a c c g a g a t g c c a t g a a t  
5761 t c c a g g g c a t c g c g t g g c g g a a a g t t c a a g g c g g a t c g t g c a c c g t t g c g  
5821 a a a t a g c c g g c g t g c a c g a c g g g c a c c g a c c g t c g g g t c c g g t t g a g c  
5881 a g g a a g a a c t c g g g a t t c a c g t g a c g a g a a g c c g a t t c c t c g c t c  
5941 a g c t g c c g c c a c a c g t g c g c g g g t c c c a c g a c g c g a g c c g c t c g g c a t g g t  
6001 a t g t c g c a a a a c a t c c g c g c t g a t t g g t g g a c c g a c c g a c c  
6061 t g g a a g g t c g a c g g g t c g g c c a c c g t a t c g g a t t c c g a g a c c c g c a a a g c c c  
6121 t c g a t c g a g g a t c c g t g a a c c g a t g c c t c g a a g g c g c g t g g  
6181 g c g a t g g c g a a c c g a c t t g a g g a a a c c g a g c g a c c a g c c g g a c g a a g c g g

6241 atgtcgctt cttccagggt acgaagaacg aattccttct gtcggtccat acctcgaaca  
6301 gtagtcactg tctgtaaaaa ccgtgttacc gatgccccgc cagaagcggt gcggggcgcc  
6361 cccgaagggg agtgcgcggt gagttcaggg cgcgcaccgc agactcgctg gcgcaaggt  
5 6421 cccgtcgaga aaatagtgc a tcaccgcaga gtccacacac tggttccat cgaacaccgc  
6481 agtgtgtgg gtgcgcgtca aggtgtatcg cggtgcgc c agctggcgcc ccaggctac  
6541 cccggactga tacggagtgg cggggtcggt ggtggtgac accacgacga cttgccgc  
6601 cccggccggc gcccgggggt gcccgcgtca cgttgcggc accggccaca gggcgcacag  
6661 atcgcggggg gcccgtccgg tgaactgccc gtagctaagg aacggggcga cctgacggat  
10 6721 ccgttggtcg gcccgcaccc aggccgtgg atcggccgg tggggcgcat cgacgcaccc  
6781 gaccgcgttg aacgcgttct ggtcggtct gtagtgc ccc tctgcattcc ggcgcgtata  
6841 gtcgtcgca agcaccagca agtcgcgc gtcgtcgcc cgcgtcagcc ccagcagacc  
6901 actggtcagg tactccagc gctgagggtc gtacagcgc ttgtatggc cgcgtcg  
6961 gtcggcgttag ctcaggccac gtggatccga cgtttaacc ggcgttgc cccagggtc  
15 7021 aaccaggggcg tggtagcggt tgaccactg ggccgagtc gtcggccagag ggcaggccgg  
7081 cgagcggcg cagtcggcg cgtagtcatt gaaagcggtc taaaatcccg ccatttgct  
7141 gatgtttcc tcgattggc taacggctgg atcgatagcg cgcgtcgagg ccattcccg  
7201 cacatgagta ccgaaccgtt ccaggtaa ggtgc cccaa tcggtgcgt agctgtatcc  
7261 gaggtatgg atctgatcgta cacttaacgc ttggcgaacc atgtccatgt cccgtcgac  
20 7321 ggacgcggta ccgatattgg ccaagaagct gaagccccatc cggtaacac acgtctggc  
7381 caactgcgg tagacgtt cgcgtgggt gacaccggcc ggactgttagt cggccatcg  
7441 atcgcgcggg tacgcgtca actcggcgac ggtgcgacac cgcacgc gggcgtgg  
7501 gcccaccct ctcgggtcg a gcccaccag gtcgaagtgg cggagaatgt cgggtcg  
7561 gatcgccggt gccatagcg g caccatgtc gaccggc gac gcccgggtc cccaggatt  
7621 gaccagcagt gtcggaaatc gtcgtccgt cgcgggacg cggatcaccg ccaacttcg  
25 7681 ttgtgtccca cccgggttgg cgtagtcgac gggacggac accgtcgcc acgtcg  
7741 gcgaatttcg ctgggtcg cgtatgcactc gggcagctg ttccaactt gttgcggcgc  
7801 cacgaccggc gcacccgggg tttggccggc gcccgggttc tcaatgcgc cggccaaacgg  
7861 gggcgtgct agggcagtc cgcgcagcag caacccgaag gacgcagc ggcagctcaa  
7921 cggctcgccg cgcacatgg cgcacatgt ctcaccggc aatacctgt acggcgaa  
30 7981 atgatcacac cttcggttct tcggcccgct agcacttggc gccgtggc ggcgtgg  
8041 cgcgcattaa atacgcgc acgtactcgta caatgcacgt gtcgcctgg aataccacc  
8101 tgtgtgggt tccgtcgaa gtcagcaacg aaccgcgaag ctgggtcgcc aggtcgaccc  
8161 cggccgttgc cggcgtcgcc gggcatgg tgggtggatc caccaccgtc ggcactaggc  
8221 cggcgcggc gacggcatgg ggctgacitg tgggtggc ac cggccagaac ggcaggtgc  
35 8281 ccagcggcgc atcaccgggt aacttccgt agctcatgaa cggcgtcgatc tcccgccgc

8341 ggccgtttc gtgcgtgacc ttgtcgatcgatcgtaaccgg gggctgatcg acgcaattga  
 8401 tcgcccacccg cgctgcacccg gaattgttgt agccggccgtg cgagtcccgac cgatgtaca  
 8461 tgtcgccag agccagcagg gtgtctccgc gatgtcgac cagctccgac agccgtcg  
 5 8521 tcaagtgttgc acacagattc ggtgagttaca ggcgcataat ggtgcccacg atggcgctgc  
 8581 tataactcg cccgcgcgaa tcctcgatgc ggcgcggcct gctgtatcgc ggggtgtccg  
 8641 ggtcgaccaa cggatcgacc aggctgttgtt agacctcgac ggcttggcc gggcggcgc  
 8701 ccagcggca gcccgcgtt ttggcgactt cggcggcata gttgttaac ggtcctgg  
 8761 agcccttggc ctggcgacgc tccgcctcgatggatcgac attgggtcg acggcaccgt  
 10 8821 cgagaatcat tgccgcacc cgctgcggaa attcctcgac atacgcggag ccgatccgg  
 8881 tgccgtacga gtatcccagg taggtcagttt tgctcgcc caacgcgcg cgaatggcat  
 8941 ccaggccctt ggcgcgtt accgtcccgatggccatggccatggccatggccatggccatgg  
 9001 ccacacagcg accgcacgaat tgcttggctt cgttctcgat gtgcgcacca ccctccgg  
 9061 tgttagtcaac ctgcggctcg gcccgcggcc ggtcggtgtc ggcacgcggatggccatgg  
 15 9121 tcgcggccgg ggacgcacgc accccgcggg ggtcgaaccc aaccaggcg aacccctcg  
 9181 gcacccgcctt cggcaatgtc tggaagacgc ccaaggccgc ctgcataccg gattgcgg  
 9241 gtccaccggg atttatgacc agcgaaccga tcttgtctcc cgtcgccggaa aagcgaatca  
 9301 ggcgcgcgc cggccacgtca ccatcgccggc ggtcgtagtc gaccggatca gcgagctgc  
 9361 cgcataacgc gccgcggggg atcttactt gccccggatggccatggccatggccatgg  
 20 9421 ccaccggctg gcccacgttgc ggctccgcatacgacgcggcc tccccggacc acgcggatgc  
 9481 agcccaacaag aaccaacgc accggccgcgac ggcgcggccca gatcaacacgc atgcgcgcga  
 9541 tcttgtcgatgc gcgagacacgc ctatgcggca atatgcgtcc agagcagacc cgagatctg  
 9601 gccagcggcc accgtcgccgactaaccgg cgcgtcccgatgc tccgcgtccatgcggatgc  
 9661 gcgaactcgatgc cggccatccc ccatacgatgc ggtaaacatgc cccggcaaga caccgacc  
 25 9721 tcgaccggat cccgcacggg cgcgtcgcc tccgggtgtc acaactgcga catcagggt  
 9781 ggcgcgtccac cccgtccacg cggcatggt gcacccgttgc catgcggccatggccatgg  
 9841 cgtgcgtc cacccttcg acgaacccat ctccacggc ggtcgccggc agcgcacgc  
 9901 tgtggccgca gatctccgatggccatggccatggccatggccatggccatggccatgg  
 9961 gcaagtgcg atcgatgtga ggttcaaggt tcagcgact gctggcaaccc ttttccgaa  
 30 10021 accgcggccct cgccttgatc tggaggatcaga acgcgtcaacg cagccgtca aaggcgtaac  
 10081 ccatgcgtca gcaaaatgc atgggctgatggccatggccatggccatggccatggccatgg  
 10141 caggccactg agccgcgtca tgcgcgtatggatgc tgcgcgtatggccatggccatgg  
 10201 gggaaagaag tggcagactg tcagggtccg acgaacccgg ggcgcgtatggccatgg  
 10261 gatgcaccccg accaccatggacatgtatgc tgcgtatggccatggccatggccatgg  
 35 10321 ccccgaggatccggcccg cggaccaagaatccgcacccatccacccatggccatggccatgg  
 10381 ccgcacgcgcaccaatgtggccatggccatggccatggccatggccatggccatggccatgg

10441 acgaggccgg catccggtg ctgctggtcg gtgattcggc ggccaacgtc gtgtacggct  
10501 acgacaccac cgtgcccgtc tccatcgacg agctgatccc gctggccgt ggcgtggtc  
10561 ggggtgcccc gcacgcactg gtcgtcgccg acctgcccgtt cggcagctac gaggcggggc  
5 10621 ccaccggccgc gtggccgccc gccacccggt tcctcaagga cggcggcgca catcggtca  
10681 agctcgaggg cggtgagcgg gtggccgagc aaatgcctg tctgaccgcg ggggcattcc  
10741 cggtgatggc acacatggc ttcacccgc aaagcgtcaa caccttggc ggcttccggg  
10801 tgcagggccg cggcgacgccc gccgaacaaa ccatcgccga cgcgatcgcc gtcgcccgaag  
10861 ccggagcgtt tgccgtcgtt atggagatgg tgcccgccg gttggccacc cagatcaccg  
10 10921 gcaagcttac cattccgacg gtcgggatcg ggcgtggcc caactgcgac ggccagggtcc  
10981 tggtaggca ggacatggcc ggggtcagcg ggcgaagac cgcccgcttc gtcaaacgggt  
11041 atgcccgtgt cgggtggtaa ctacgcccgtg ctgcaatgca atacgccaa gaggtggccg  
11101 gcggggtatt ccccgctgac gaacacagt tctgaccaag cggatcagc cggatcgccg  
11161 ggcattgcgg tggccctgtg atgcccgtg acgcccggatt gccggccgg acgcgccagc  
15 11221 gggacccatc ggcgtcgctg tcgcccggatg agcccggtt gagcccagac attcgatgt  
11281 cccaaacacca tccgcccacag cccaaattgtat gtggcactct atgcatgcct atcccccacc  
11341 aaccaccacc gccggcgacgc atcatgaccg gaggcgaaga tgccagtaga ggcgcccaga  
11401 ccagcgcgcc atctggaggt cggcgcaag ttcgacgtga tcgagtcgac ggtgtcgccg  
11461 tcggtcgagg gcatcgccgc ggtgggtcgc gtcgagcgtt cggcggccca cggatcgac  
20 11521 gcggtgtact tcgacacacc gtcgacacgac ctggcgcgc accagatcac ttggcggcgc  
11581 cgcaccggcg ggcggacgc cggctggcat ctgaagctgc cggccggacc cggacacgc  
11641 accgagatgc gggccacgt gtcggcatca ggcgacgcgt tgccggccga gttgtggat  
11701 gtgggtcgcc cgatcgccgc gggccacgt gtcggccgtt tcgcccggat cggcgtcc  
11761 cggcggccgg agatcctgta cggcggccgg ggcgacgcgc tggcggaaatt ctgcaacgc  
25 11821 gacgtcaccg catggcggc cggggcatttc caccggcgctg gtgcagcggaa caacggccct  
11881 ggcggccggc agtggcgcga atgggaactt gaaactggtca ccacggatgg gaccggccat  
11941 accaagctac tggaccggctt agccaaaccgg ctgctcgatg cgggtggccgc acctggccgc  
12001 cacggctcca aactggcgcg ggtggccgtt ggcggccgtt ccgggtggatc gcccacggc  
12061 cggcggccgc cggccggatcc agtacaccgc ggggtgtcccg agcaagtcga cggatcgctg  
30 12121 ctgtgggatc gggccgtcg gggccggcc tatgacgcgc tgccggatc gggatcgac  
12181 acccgcaaga tcccgccgtt gtcgacggat tcccgccgtt cgttggccctt gaaggaaat  
12241 ggcgtgggtca tcgatgaact ggcgtggatc ggcgtggatc tggcggccgtt cggccggcc  
12301 gaggtactcg gtgaccgcgtt ccacggcgaa ctggacgcgc tggcggccgg gctggatcg  
12361 ggcgtgggtca ggcgtggccgtt ggtggccgtt ggcgtggccgc gataccagac cggccggcc  
12421 cggatcgatcactga tcgcatcgatc gtcgacggatc tccatcgatc tccatcgatc  
35 12481 ctgtgtccg aacgcggccca tgccacttctt ggggaggaat cggcaccggat aaccatcgat

12541 gcggcctacc ggcgagtccg caaagccgca aaagccgaa agaccgcgg cgaccaggcg  
12601 ggcgaccacc accgcgacga ggcattcac ctgatccgca agcgcgaa gcgattacgc  
12661 tacaccgcgg cggctactgg ggccgacaat gtgtcacaag aagccaaggt catccagacg  
5 12721 ttgttaggcg atcatcaaga cagcgtggc agccggaaac atctgatcca gcaggccata  
12781 gccgcgaaca cccgcggcgaa ggacacccttc acctacggc tgctctacca acaggaagcc  
12841 gacttggccg agcgctgccc ggagcagctt gaagccgcgc tgccaaact cgacaaggcg  
12901 gtccgcaaag cacgggattt agccgcggac gggccggacga tttggctgt aagccggatt  
12961 ctgtccgcg cccgcacagc caagctaacc gcggcacggc ggcgaccatc catctggaca  
10 13021 caccgttacc ggggcctcg agcggctac ccgcaggctc gggcgagcaa ccctcaagcg  
13081 cctgcgcggc cgcacttgc tgccgcctt ctggccctt cttcggtgg ggtttgccta  
13141 gccaccccg tcacccggaa tgctggcgc ctcttaccgc accgtttcac cttggcacc  
13201 acgaggatgg cggctgttt tctgtggcac ttcccgca gtcacccggg attggcgta  
13261 gcaatcaccc tgctctgtga agtccggact ttccctgact cgacgctgaa cctcgtgaat  
15 13321 ccacacaagc cctacgcgag ccgcggccgc ccagccaaact catccgcac gaccacgcta  
13381 ccccgctggg cgggtgcgcg gccagtgtga ccgcgtggacg acacggctag tcggacagcc  
13441 gatccggcgg gcagtcctta tcgtggactg gtgacacggt gggacaaacg cgtcgactcc  
13501 ggcgactggg acgcctatgc tgccgaggc agcgagtagc gtggcgcact gctacccgg  
13561 ctgatcaccc cccgcgaggc cgcggctg cgcaagctgt acgcccacga cggcctgtt  
20 13621 cgctcgacgg tgcataatggc atccaagcg tacggccgg ggcagtatcg atattccat  
13681 gccccctatc ccgagtgtac gagegtctca agcaggcgct gtatccaaa ctgctggcga  
13741 tagcgccaa ctggtggcc aaactgggccc gggaggcgcc ctggccagac agccttgatg  
13801 actgggtggc gagctgtcat gccgcggcc aaacccgatc cacagcgctg atgttgaagt  
13861 acggcaccaa cgactggAAC gcccataacc aggtatctca cggcgagttt gtgttccgc  
25 13921 tgcagggtgt gatcaacctg agcgatccgg aaacccgacta caccggccgc gaggctcg  
13981 ttgtcgaaca gccgcctcgcc gccaatccc ggggtaccgc aatgcaactt cccgcaggac  
14041 atggttatgt gttcacgacc cgtgatccggc cggtgccggac tagccgtggc tggcggcat  
14101 ctccagtgcg ccatggcctt tcgactattt gttccggcga acgctatgcc atggggctga  
14161 ttttcacga cgcagcctga ttgcacgcca tctatagata gcctgtctga ttccaccaatc  
30 14221 gcacccgacga tgcccatcg gcgtagaact cggcgatgtc cagcgatgcc agatcaagat  
14281 gcaaccgata taggacgccc gacccggcat ccaacgcgcag ccgcacaaac attttgcgt  
14341 gcgtgacatg tgacaccacc agcaccgtcg cgccttcgtc gccaacgatc atccgatcac  
14401 gtcccccgg aacccggccgc agcacgtcg cgaagcttc cccacccggg ggcgtgtac  
14461 tgggtcctg cagccagcga cggtgccggc cggatccggc ttctgcggcc tccgcgaacg  
35 14521 tcagccccctc ccaggcgccg aagtccgtct cgcaccaggc gtcacgacg accacgtcca  
14581 gggccaggcgc tctggccggc gtcacccggc tgcgttaacg cccgtgtac ggcgaggaga

14641 ccacccgcac gatcccgccg cgccgcgcca gatacccgcc cgccgcacca acctggcgcc  
 14701 accccaccc tcgtcaacccc gggttgcccgc gccccgaata gcggcggtgc tccgacagct  
 14761 ccgtctgccc gtggcgcaac aaaagttagtc gggtgggtgt accgcgggccc cccgtccagc  
 5 14821 cgggagatgt cggtgactcg gtgcacacga ttttgcagg atccgcattcc gccgcagccg  
 14881 attgcgcggc ggcgtccatc ggtgtcaatgg ccaacccggc tgcatacgtg ttccgggcac  
 14941 gcggaaacca ctcgtatgt atccgtcgaa actgggacgc caacgcctga gcctggacat  
 15001 agagttcag cagatccggg tgcttgaccc tccaccggcc ggacatctgc tccaccacca  
 15061 gcttggagtc catcagcacc gcggcctcg tggcacctag ttacacggcg tcgtccaaac  
 10 15121 cggctatcag gccgcggat tccggcgtacgt tggtcgctgc cccggccgatc gcctgcgttgg  
 15181 actcggccag cacgggtggag tgatcgccgc tccacaccac cggccgtat cccggccggc  
 15241 cgggattgcc cccgcgtcccg ccgtccgtt cgtacacaaatcccttc  
 15301 gagccgcaac aagatcgctc cgttccgg gcagcgcacc acttcaccc cggcggccgc  
 15361 cggatctgg gccagtcgc cccggccgat ctcgtccgg caggcaccac atcgatgacc  
 15 15421 ttgcacccgc cccggccctg gcccgcctcc ggcccgctgt ctgttgcgttgc gccccgcaag  
 15481 ctcggatca agtgcgcgc tcagcatgtc gctgtcgat gaatgttggt gccgggcttgc  
 15541 gtcgatttcg gcaagtgcct cgtccaaagc ctgtggcgcc gcccgcagg cggcccgca  
 15601 cgttggagc gcccgcact cggcggtctg ttgagccctgc agtccctcg ggcgttccag  
 15661 caccccccac agggcatctt ccaaactggc ttgacggcg tgcacgtgt cgagctcg  
 20 15721 ctgcagatca gccaattgtc tggcgtccgt tgcacccgaa gtgagcaacg accggcccc  
 15781 gtcgccacgc ttacgcaccc catcgatctc cgactaaaaa cggcacaccc gcccgtccaa  
 15841 gtcctccgc gcatcgca gggccgcatt cctgtcggtt gcccgcgttgc gtcggcctg  
 15901 caccgtctgg taagccgc gctgcggcag atgggttagcc cgtgcgcga tccgggtc  
 15961 ctcagcatcc agtgcgcctt attccagtag cggccgttgc tgcacccactc cggcttcat  
 16021 gcctgatctc tcccaatttc gtgtcgagg ttccacgggt cgggtcgat ggtgcacaca  
 25 16081 cgcacccggca ggcacgcgc gaaatgagac cggcacaccc cggccgcctg gcccaccc  
 16141 gggaaattcgca ttggccaaatg cggcgtcgatccatggcc ctgtcgaaatc tggcaatgc  
 16201 tcgtcggtctg gatgtcgatcg cggatcgcc gtaacgtacg ctgtcgatcg cggcgcc  
 16261 acggtggcaa gcaacgagtc cccggcgcccg cggcagaccc cggccgcga caccagc  
 16321 tcggatcccc cggcgccgcg cacaccggtc gcggatcgcc gcaacgcggc ctccagacgg  
 16381 gcaacaaagg tgcgcagccgg ttccgggttt ggcgtctgc caatccggcc taaccgc  
 16441 cccgcacccggc gttgtaccatcg cggcgtatcg tcgtatcgcc gtcgtcgatcc  
 16501 ggcgcacatcg cccgcaccc ctcggcgccg gtcgtcgcc gtcgtcgatcc ctcgc  
 16561 tccgtcgccca cccgttcgac ggtaccgacg ctgtcgatcc gggcgacgc cccgtcg  
 16621 gccaggaact gcccggtacc cggcgtatcg cggatcgact gtcgtcgatcc  
 30 16681 cccgcacccgg cctcaaaagac ctcgtccgc accgcctcg agtgcgcgc cggcacatcg  
 16681 cccgcacccgg cctcaaaagac ctcgtccgc accgcctcg agtgcgcgc cggcacatcg

16741 atgaccact tgcgagatc ggccgctccg ggcaccgggt cgagaacggc gtcgacggtc  
 16801 agaccaacag cgtgtgccag cgcgtccgac acacccggcg acgcccggc ggcgtggtg  
 16861 tgcgcggtaa acaacgagcg accggtccgg atcaggcggt gcaccagcac acccttggc  
 5 16921 gtgtggccg cgaccgtatc gacccacgc agtaacaacg ggtggtcac caatagcagt  
 16981 ccggcctggg gaacctggtc caccaccggc ggcgtcgct ccacccgaac ggtcaccgaa  
 17041 tccaccacgt cgtcggggtc gccgcacacc agacccacccg aatccacga ctggcaagc  
 17101 cgccgggggtt aggcctggtc cagcacgtcg atgacatcg ccagccgcac actcatggc  
 17161 gtcctccacg ctggccac tcggcgatcg ccggcaccag cacggccac tccggcgca  
 10 17221 ccggccggccg caggtaccgc gcgtccaggc cgacgaaggt gtcaccggc cgcaccgcaa  
 17281 ttcccttgct ctgaaatag ttctgtaatc cgtcagcatc ggcgtatgtt aacagtacga  
 17341 aaggggccgc accatcgacc acctcgaccc ccaccgtatc cagtccggcc accatctccg  
 17401 cgccgcgcgc cgtcaaccgc accgcattcg ctgcggcagc ggcgaccgc cggggggcgc  
 17461 agcaagcagc gatggccgtc agttgcaatg ttcccaacgg ccagtgcgtc cgctgcacgg  
 15 17521 tcaaccgagc cagcacgtct ggcgagccg ggcgttagcc cacccgaat ccggccagcg  
 17581 accacgtttt cgtcaagcta cggagcacca gcacatcggtt cagcgagtca tcggccaacg  
 17641 attgcggctc gccgggaacc caatcagcgac acgcctcgatc gaccaccagg atgcgtcccg  
 17701 gccggcgtaa ctcgagcgc tgctcggtt ggtgcagcac cgaggtgggg ttggcggtt  
 17761 taccacacgac gacaagggtcg gcgtcgtagt gcacgtgcgc ggtgtccagc acgaacggcg  
 20 17821 gctttaggac aacatggtgc gccgtgattt cggcagcgct caaggctatg gccggctcg  
 17881 tgaacgcggg cacgacgtt gctgcccgc cggacttagt gttgtgcagc aatgcgaatc  
 17941 cctccggccgc cccgacgagc gggagcactt cgtcacgggt tctgcatgc cgttcagcg  
 18001 ccgcgtcttgc cggccgggtgc acatcgatc tgctcggtt gccggccagc tccggcagca  
 18061 ggcggcggag ctgcccggacc aaccattccg gggccgggtc atggccggacg ttgacggcg  
 25 18121 agtccagcac gcccggcgac acatcctgtt caccgtggta ggcggccgc gcaagcgggc  
 18181 tagtgtctat actcgccaca gcgtcaaaca gtgtggcc ggtgtcggtt ccaagaatcc  
 18241 agagcaccgc cgcgcgttgc tctacgcggc gacaaccgcg acatcacagg cagctaacag  
 18301 ggcgtcgccg gtgtatcg tcaaggccaaag cagctgtgcc tggcgatga gcacacggc  
 18361 gaatggatgt cgtatgttgc tccggaaatc tgcggcgcc ggtgtgtcg tggtaactg  
 18421 acagcggcga cgtccgcag cggcgcattt gatcgggcac gtaagaagcc gatggctcg  
 30 18481 ggcggccggag ctggccggagg cggtagttgc tgctcgatctt ccaggactg gccggccgaca  
 18541 agagaatgtt gttgcggacg tccgtaaatc tgcggcgcc ggtgtgtcg tggtaactg  
 18601 ccaaactgttgc gttgcgtatgc ggtgcgtt caccgtggaa agcgttcgag cacgtcgatc  
 18661 gacaacggag cgtccaaatc gtcggccacg cggtaacacgc catggtaat gcctaaccgc  
 18721 cgagtcgtatc gaggatgcag cggcacaacgc ttgcgttccgc gtcggccgcg gcccggaaatc  
 35 18781 tcaacctctg cccgcccgttgc acgagccgc gcaagctcgatc caggcgtgtc ttgcctcg

18841 gaacgcccac ccgcttcgca ggcgcccaga ctttcgcgtc gaccacctgc tcaccaaact  
 18901 tcgcgtatcat cgccgtatac cacagcgcca acgggttagcg gtttgcctaa ccgcttcgat  
 18961 aacgacaatg ggatcgtgac cgacacgacc gcgagcggga ccaattgccc gcctccat  
 5 19021 cgccgcgcgcg cacggcgcgc atcgtcgccg ggtgaatcgc cgcaatgggt gatctcgat  
 19081 ctggacggca cgctgaccga ctggcgccgc ggaatcgat ccagcttccg acacgcgc  
 19141 aaccacatcg gtccccactg acccgaaaggc gacctggcca ctcacatcgat cggccgc  
 19201 atgcatgaga cgctgcccgc catggggctc ggcgaatccg cggaggaggc gatcgtagcc  
 19261 taccggccgc actacagcgc cccgggttgg gcgatgaaca gtttgcgtca cggatcg  
 10 19321 ccgctgctgg ccgacactgcg caccggcggt gtccggctgg ccgtcgccac ctcaaggca  
 19381 gagccgaccg cacggcgaat cctgcgcac ttggaaattt agcagcactt cgaggctatc  
 19441 gcgggcgcga gcacccgttgg ctgcgcaggc agcaaggctg acgtcgatcc ccacgcgc  
 19501 ggcgcgcgcg ggccgcgtacc cgagcggttgc gtgtatggcgc ggcgcgcgc ccacgcgc  
 19561 gacggggcgg cccgcacccg catcgacacg gtgggttgc gctggggcta cggccgcgc  
 15 19621 gacttatcg acaagaccc caccaccgtc gtgacgcattt ccgcacatg tgacgagctg  
 19681 agggaggcgc taggtgtctg atccgctgca cgtcacattt gtttgcgtt gcaacatctg  
 19741 ccggcgccca atggccgaga agatgttcgc ccaacagctt cgccaccgtt gcctgggtga  
 19801 cgccgttgcga gtgaccatgtt cgggcacccgg gaactggcat tggcgttgc ggcgcacga  
 19861 gcgggcggcc ggggtgttgc gagccacccg ctaccctacc gaccaccggg ccgcacaagt  
 19921 cggcaccgaa caccggcggtt cggccatgtt gttggccatg gaccgcaccc acgctcg  
 20 19981 gttggccatgtt ctcggcgatcc aagccgcgg ggtacggatgtt ctgcggatcc tcgacccac  
 20041 ctcggaaacc catgcgcgttgc atgtcgatcc tccctactat ggcgtatccact ccgacttcga  
 20101 ggaggcttc gccgtatcg aatccgcctt gcccggctt cacgactggg tcgacgaac  
 20161 tctcgccgg aacggaccga gttgatgcgc cgcctagctt tcctgttgc gcccggctt  
 25 20221 ctggcggttgg ccctgggttgcgtt ggtcggttc acctacccgtt gtttacgggt gtcgcgc  
 20281 tggcagctgg gcaagaatgc caaaacgtca cgagagaacc agcagatcgat gtattccctc  
 20341 gacacccgcg cgggtccgtt gaaaacccctt ctaccacagc aggattcgatcc ggcgcgc  
 20401 ggcgcgttgc gcccgggttgc ggcacccggc cagtacccatcc cggacgttgc ggtgtgg  
 20461 cgactgcgcg tgggggggggg ggaccaggcg tttggggatgtt tggccctt cgtggatcg  
 30 20521 ggcggacca cccgccttgcgtt ccgcgttgc gtttacccgtt gtcgcacgtt  
 20581 ccaccgttgc cccgccttgcgtt ggtcgatcc cccgcgttgcgtt ggcgtactcc  
 20641 gaaccgagcg tggggggccaa agacccattt gtcagagacg gtttccatcc ggtgtattcg  
 20701 atcaataccg gacaggttgcgc cccgcgttgc ggttccatcc tggctgggtt cttatcg  
 20761 ttgatcgatcc accaaccggc cgggtccggc gtttgcgttgc ttccgttgc ttttccatcc  
 20821 ccgttccgtt ccgttgcgtt ccgttgcgtt ccgttgcgtt ccgttgcgtt  
 35 20881 ggctatttcg cccacccggc ggggtccggc gtttgcgttgc ttccgttgc ttttccatcc  
 20881 ggctatttcg cccacccggc ggggtccggc gtttgcgttgc ttccgttgc ttttccatcc  
 20881 ggctatttcg cccacccggc ggggtccggc gtttgcgttgc ttccgttgc ttttccatcc

20941 ccggacaaggc caatgacggt cgagcagaaa ctcgctgacc gctacggccg cggcggtaa  
 21001 accaacatca cggccaatac cgcagcccc gcctggacca cccgcacag caccacggcg  
 21061 cggcgacat cggccaccc gggcgaccgg ccgtcgccca aggtggccg gatctgcaac  
 5 21121 tcatggtggt accgggtggg cccacccagc cgcacgtcaa gcgcggcagc aaacggccgc  
 21181 tcgacgacac cggcggtgg gctggatgg cggcgccgt cgcggccca ggcccggtacc  
 21241 gcaccgcggg gcgacccacc gaccacggc ggcgagatca ccaccagcac cgccgtcgcc  
 21301 cgtgcgccaa catagttggc ccagtcattc aatctgtctg cagccaaacc gaatcgag  
 21361 taacgcggcg agcggttagcc gatcatcgag tccagggtgt tgcgtgcacg atatccac  
 10 21421 accgcaggca cgccgctcgaa agccgcccac agcagcggca ccacctggc gtcggcggtg  
 21481 tttcgccca ccgactccag cgcggcacgc gtcaggcccg ggccgcccag ctggccggg  
 21541 tcacgcccgc acagcgacgg cagcagccgt cgcggccct cgacatcgac gcgcgtccaaac  
 21601 aggtcgata tctggggcc ggtgcgcgc agcgaagttc cgcccgacgc tgcccggtg  
 21661 gccgtcgccgg tggccggccac gggccaggac ctggccggta gccgtcgac tgcccgccg  
 15 21721 agcaagccca ccgcgcggac cagcaggccg acgttaccg caccggcgac cggccgtca  
 21781 cggtaggtga tctgtccag cttggggcc gcccggccaa acaggccac cggatgaccc  
 21841 cgttgggggt cggcgaacac gacgtcgacg aggccggccaa tcagcagccgac gacggccctg  
 21901 gtcgtccagg tcgatcaaa cactccggca gcgtcgacaca cgtggcttac gtcagctat  
 21961 ttatgacccatc atacggcagc tatccacgtt gaagcggcca gtcacccggg ttggccgaccc  
 20 22021 gttgaaccccg gccggcaatgt tggccggcc acgcaatgtc atcatgcacg tggcaatgtcc  
 22081 ggggtcggttatggcggtgc tggaaagccc ggtggacacgc ggcaacgtct acaagcatcc  
 22141 gttcaagcgg gcccggacca cggcaccta cttggccggta ggcgaccatcg ggacggaaatc  
 22201 cgaccgagcg ctgtccggg tggccgtgg cgtcgccac cggcagggttc gtcgacggc  
 22261 ctgcagccca gtgtcctata acgccttcga cccgaagtgg cagctgtggg tggccgggtg  
 22321 tctgtaccgc tacttcgtgg accagcacga gtttctgtac ggcggccactcg aagatgcac  
 25 22381 cggccgacgccc gtctaccaag acggccaaacg gtttagggacc acgtcgagg tgccggagg  
 22441 gatgtggcccg cgggaccggg tcgcgttcga cgagtactgg aagcgtcgac ttgtatggct  
 22501 gcagatcgac ggcgggtgc gcgagcatct tcgggggtt gcctcggtt cgtttctccc  
 22561 gtggccgttg cgcgcgggtgg cggggccgtt caacctgttt ggcgacacgg gattcttggc  
 22621 accggagttc cgcgcgtatca tgcagctggta gttgtcacag gcccggcagc gtcgttcga  
 22681 gtgggttactt tccgtgtac gtttagccga cggcgttattt ccgcacatggg cctggatctt  
 22741 cgtttaccag ctttacttgtt gggacatgcg gtttcgcgc cgcacacggc gccgaatgt  
 22801 ctgatagagc cggcccgagt gtgagccgtaa cagccggaca cggcggcgt gtgtcggtc  
 22861 gcaagggttca cgcgcggcgtaa tctagagccg cggaaaacctt acitctgggt tgcctccg  
 30 22921 atcaacgttc tgcgttcgc ggcacatcgca cgcacatcg gtcgttcgc atccaccgc  
 22981 gcatacaggt cggccgtggt cggccggcagc tggccgacg tcattggccg caccggcggt  
 35

23041 gctgtctgtc gcgccgcgtc gtgcgttga aacccaggc gtcacccac gaccacgaca  
 23101 ctgcctatc cggcgccccg ccgacaacga agcacagcta gccgggtggc gcggacggga  
 23161 tcgaaccgccc gaccgttgtt gtgtaaaacc agagctctac cgctgagcta cgcgcccatt  
 5 23221 accgcccgcag gctacacgccc ttgcggccaa gcacccaaa ctttaggccc taagcgccgc  
 23281 cagagcgtcg gtccacagcc gctgtatcg aacttcaccc ggctgttca tctcggcgaa  
 23341 ccgaatgtatc cctgaccgat ccgaccacaaa ggtgccccgg tttagcgatgc cggcctgctc  
 23401 gtgtaaagacg ccgtaggcct gactgaccgc gccgtgtggc cagaagtccg acaacagcgg  
 23461 aaacgtaat ccgtctgctc tcgcccagat ctgtgagtg ggtggcgggc ccaccgaaat  
 10 23521 cgctagcgcg ggcgtgtcgt cgttctcaaa ctgcggcagg tgcacgcgca actggccag  
 23581 ctgcctctgg cagatgccc tgaacgcca cggaaagaac accaacagca cgttcttgc  
 23641 accccggtag ccgcgcagg tgacaagctg ctgattctgg tcgcgcaacg tgaagtcagg  
 23701 ggccgtggct ccgcgcgttca gcatcagcgc ttgcagccc gcgattcgg ctgtaccaat  
 23761 ctgctggcgc tccagttgcc cagattgacc gacgaggctg gcatcagccc agctgtggc  
 15 23821 gccgcctcgg caatctcggc gggcaataca tggccggctt ggccggctt gggcgtacc  
 23881 acccaaatac caccgtctc ggcgagcggg ccgatcgcat ccatcagggt gtccaccaaa  
 23941 tcgcccgtcgc catcagcaca ccacaacagg acgacatcgatc tgacctcg tgcgttca  
 24001 tcgagcaact ctccccgc cgcgttcgc atggccgc ggtgtcg tgcgtgtct  
 24061 tcgtcccgac cccattccgt gataagttgg ttcgttggta tgcccaattt gccccgttag  
 20 24121 ttcgaggcgt gatccgcgc gaccaccgtg gaaacctt cagtcgcgc gggccatgt  
 24181 cacaccgtcg cgtatggcat tatcgatcgca cagccagaac cggccaccc gcccgcctca  
 24241 gaaggccggcc acgcacattg tcaatgcctt tgcgttggc tgcgttggc gatcaacccg  
 24301 ccgggtgaat tccgcgtcg acgcgtgcgc accgatggca ttgcaccc cgccggccgc  
 24361 gtgcacatat gcggttgcgc catccccag ttgcgcggac agcgcggcgc tcagactgcc  
 25 24421 tgagaccgtc gaggcactgt tgcgttggc ggcgttccgg tggccgggt  
 24481 gttggccccc tgattgaacg cggccacgtt ggcgttccacc ttgcgttggc cgttccgt  
 24541 ggtggccgc acgcgtcac acgagggtcg aatgcctt gtcgttgcg attgttggc  
 24601 ctgcgactcc cggatgtcg acgtcgccgc cgaagccgac accgacgcgg acaccgacga  
 24661 gcggttggcc ggtgcgttgc tgggttgcggc catggccgtt ccgttgcgttga cagtggtaca  
 24721 tccgacgtatcc cccatcagca gcagcgcgtt gcaaggccgc ctcgcctgg  
 30 24781 gagctcccccc cctgttgcgc gaggcaccgc ggcgcacccatccg atgaggacccg catgttgg  
 24841 tacctggcgtc cagcgcgacc ggcgttgcgc tgggttgcgtc cgcacccatccg  
 24901 gaggttgcgttgc atccgcgc gacgcgttgc cggcacgata gggggacccg catctaaaca  
 24961 gcacgcgttgc ggaagccgc cacctacagg agtagtgcgt tgaccacccgatccgc  
 25021 cacgatctgg cccaaactc aaacagcgttgc agcgttgc accgaggatcg ggtgtatccgc  
 35 25081 gaggttgcgttgc cgttgcgttgc gcccgttgc gacccgttgc gtcgttgc  
 25141 gaggttgcgttgc cgttgcgttgc gcccgttgc gacccgttgc gtcgttgc

25141 tcctttgaca cgctgctgca acgctgcggc ccgtcgcgaa cccgctacct gatgttgcgg  
25201 ctgcttagagc gggccggcga gcagcgggtg gccatcccg cattgacgac taccgactat  
25261 gtcaacacca tcccgaccga gctggagccg tgggtcccg gcgacgaaga cgtcgAACgt  
5 25321 cgttatcgag cgtggatcgat atggaatcgcc gccatcatgg tgcaccgtgc gcaacgaccg  
25381 ggtgtggcg tgggtggcca tatctcgacc tacgcgtcgat ccgcggcgat ctatgaggtc  
25441 ggttcaacc acttctccg cggcaagtgc caccggggcg gcccgcgtatca ggtttcatc  
25501 cagggccacg ctccccggg aatctacgcg cgcgcctcc tcgaaggcg gttgaccgcc  
25561 gagcaactcg acggattccg ccaggaacac agccatgtcg gcccgggtt gccccttat  
10 25621 ccgcacccgc ggctcatgcc cgacttctgg gaattccca ccgtgtcgat gggtttggc  
25681 ccgcacccgc ccacatccca ggcacgggtc aaccactatc tgcatgaccg cggtatcaaa  
25741 gacaccccg atcaacacgt gtgggtttt ttgggcgacg gcgagatgga cgaacccgag  
25801 agccgtggc tggcccacgt cggcgcgtg gaaggcttg acaacttgac ctctgtgatc  
25861 aactgcaatc tgcaacgt cgacggcccg gtgcgcggca acggcaagat catccaggag  
15 25921 ctggagtcgt tctccgcgg tgccggctgg aacgtcatca aggtgggttg gggccgcgaa  
25981 tggatgccc tgctgcacgc cgaccgcgac ggtgcgtgg tgaatttaat gaatacaaca  
26041 cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccactc  
26101 ttcggccgcg acccacgcac caaggcgctg gtggagaaca tgagcgacca ggatactgg  
26161 aacctcaaac gggcgccca cgattaccgc aaggttacg ccgcctaccg cgccgcgc  
20 26221 gaccacaagg gacagccgac ggtgatcctg gccaagacca tcaaaggcta cgcgcgtggc  
26281 aagcatttcg aaggacgca tgccacccac cagatgaaaa aactgaccct ggaagacctt  
26341 aaggagttc gtgacacgcg cgggattccg gtcagcgacg cccagcttga agagaatccg  
26401 tacctgcgc cctactacca ccccgccctc aacgccccgg agattcgta catgctcgac  
26461 cggcgccggg ccctcgaaaaa cttgttccc gagcgcagga ccaagtccaa agcgtgacc  
25 26521 ctgcgggtc ggcacatcta cgccgcgtg aaaaagggtt ctggcacca ggaggtggcc  
26581 accaccatgg cgacgggtcg cacgttcaaa gaagtgggtgc ggcacaagca gatcgccgc  
26641 cggatagtcc cgatcattcc cgacgaggcc cgcacccgttggactc ctgggttcccg  
26701 tcgctaaaga tctataaccg caatggccag ctgtataccg cgggttgcgc cgcacccgt  
26761 ctggcctaca aggagagcga agtcggcag atcctgcacg agggcatcaa cgaagccggg  
30 26821 tcggtggct cgttcatcgcc ggccggcacc tcgtatcgca cgcacaacga accgatgatc  
26881 cccatitaca tcttctactc gatgttgcgc ttccagcgca cggcgatag ctctggcc  
26941 gcggccgacc agatggctcg aggggttgcgtg ctggggccca cgcggggcg caccaccctg  
27001 accgggtgagg gcctgcaaca cggcgcacggt cactcggtgc tgctggccgc caccaccctg  
27061 gcgggtgggtg cctacgaccc ggcctcgcc tacgaaatcg cctacatgtt ggaaagcgga  
35 27121 ctggccagga tttgtggggaa gaacccggag aacatcttct tctacatcac cgttacaaac  
27181 gagccgtacg tgcaacgttccg cggagccggag aacttcgatc cgcaggcgatc gtcgcgggt

27241 atctaccgct atcacgcggc caccgagcaa cgcaccaaca aggccgagat cctggcctcc  
27301 gggtagcga tgcccgccgc gctgcggca gcacagatgc tggccgcccgtatggatgtc  
27361 gcccggacg tgtggcggt gaccagttgg ggcgagctaa accgcgacgg ggtggccatc  
5 27421 gagaccgaga agctccgcca ccccgatcgg cggcgccggc tgcctacgt gacgagagcg  
27481 ctggagaatg ctcggggccc ggtgatcgcg gtgtcgact ggatgcgcgc ggtccccgag  
27541 cagatccgac cgtgggtgcc gggcacatac ctcacgttgg gcaccgacgg gttcggttt  
27601 tccgacactc ggcccgccgc tcgcccgtac ttcaacacgg acgcccatac ccagggtggc  
27661 gcggtttgg aggcggtggc gggcgacggc gagatcgacc catcggtgcc ggtcgccggc  
10 27721 gcccggccagt accggatcga cgacgtggcg gctgcggcccg agcagaccac ggatccgggt  
27781 cccggggcct aacgcggcg agccgaccgc ctttggccga atctccaga aatctggcgat  
27841 agcttttagg agtgaacgac aatcagttgg ctccagttgc cggcccgagg tcgcccgtcg  
27901 aactgtggaa cactgtgcc gattcgctgc tggcgccgtt gaagcgtac tcggccggc  
27961 tggccaccga ggcgtttcg gccatgcaag aacgggtgcc gttctcgcc gacctagaag  
15 28021 cgtcccagcg cgccagcgtg ggcgtgggg tgccagacggc cgtggtaac ttcggtcaat  
28081 ggatgcacga cccgcacagt gacgtcggt ataccgcga ggcattcgag ctggtgcccc  
28141 agatctgac ggcacggatc ggcgtgcgc agaccgtggc catgggtgcgg gtcaccatgg  
28201 agttttcga agaagtcgtg cccctgctcg cccgttccga agagcgttgc accgcctca  
28261 cggtggcat ttgaaatac agccgcgacc tggcattcac cgccgcacg gcctacgcgg  
20 28321 atgcggccga ggcacgaggc acctgggaca gccggatggc ggcacgcgtg gtggacgcgg  
28381 tggtaacgcgg cgacaccggc cccgagctgc tggccggc ggcgcgcgtg aattgggaca  
28441 ccacccgcgc ggcgaccgtt cttggggaa ctccggccgc cggtccaaat ggctccaaca  
28501 ggcacggcga cagcgacggc gccagccagg atgtccgcga caccggcgtcgccacggcc  
28561 ggcgtgcgt gaccgcgtg cacggcacct ggcgtggc gatcgctcc ggcacgcgt  
25 28621 cgccaaaccga gaagttccctc aaagacctgc tggcagcatt cgccgcacgc cccgtggta  
28681 tcggccccac ggcccccattt ctgaccgcgg cgccacccgcg cgctagcgag ggcacgcgt  
28741 ggatgaacgc cgtgcggc tggcgccggc ggcgcgcgtt cttgggttggcgtt agggactt  
28801 tggccgaacgc ggcgcgtatg ggcgcgcgtt cggcgtatgtt ggcgcgttcat accgcgttga  
28861 tggcccccctt agccgatgcc ggaccgcgc tcatcgagac gctagacgc tatctggatt  
28921 gtggcgccgc gattgaagct tggccagaa agttgttcgt tcatccaaac acagtgcgg  
30 28981 accggctaa gggatcacc gacccatcg ggcgcgtatcc caccgcacgc cgcgtatcc  
29041 atgtccctcg ggtggccggcc accgtgggtc aactcaacta tccgacgcgc cactgaagca  
29101 tcgcacacaa tggccgttca tagattccctt cggcgttgcg agggggccca gcaggggccc  
29161 cggaaagata ccaggggcgc cgtggacgg aaagtgtatcc agacaacagg tcgcgggacg  
29221 atctaaaaaa catagcttac aggccgtt tttttttat atacaaaaac ctaagacgag  
35 29281 gttcataatc ttttacaccgc cgccaaaccgc ttttcacagt gttcttttagt acacgtgtt

29341 gcgttgctcg cacccggaca gggttcgcaa accgaggaa tggcgccgttgcg 29401 ctgccccggcg cagcggacca gatcgccggcg tggcgaaag ccgctgatct agatcttgc 29461 cggctggca ccaccgcctc gaccgaggag atcaccgaca ccgcggctgc ccagccatg 29521 atcgcccg cgactctgtc ggcgcaccag gaactggcgcc gccgatgcgt gctgcggc 29581 aaggacgtca tcgtggccgg ccactccgtc ggcgaaatcg cggctacgc aatgcgggt 29641 gtgatagccg ccgacgacgc cgtcgccgt gccgcaccc gggcgccga gatggccaa 29701 gcctgcgcca ccgagccac cggcatgtc gcggtgtcg gggcgacga gaccgagggt 29761 ctgatcgcc tcgagcagct cgactggc cggcaaaacc gcaacgcgc cggccagatc 29821 gtcgtgcgc gccggctgac cgccgtggag aagctgcgc aagaccgc gccaaggcg 29881 cgggtgegtg caetgggtgt cgccggageg ttccacaccc agtcatggc gcccgcactt 29941 gacggcttg cggccgcgc ggcacacatc gcaaccgcg accccaccgc cacgctgtcg 30001 tccaaccgcg acgggaagcc ggtgacatcc gggccgcgg cgatggacac cctggctcc 30061 cagtcaccc aaccgggtcg atggacactg tgcaccgcg acacacagtc 30121 acggcgtcg tggagttccc ccccgccggc acgcgtcg gtatgccaa acgcgaactt 30181 cgggggggttc cggcacgcg cgtcaagtc cccgcagacc tggacgagct ggcacacacta 30241 taaccgcgga ctggccaga acaaccacat acccgtcagt tcgattgtcacaacat 30301 tacgaaggga agcatgtgt gcgtgtact caggaagaaa tcaattgcgg tatgcggag 30361 atcatcgaag aggttaaccgg tatcgagccg tccgagatca cccggagaa gtcgttgc 30421 gacgacctgg acatcgactc gctgtcgatg gtcgagatcg ccgtgcagac cgaggacaag 30481 tacggcgtca agatccccga cgaggacctc gccggcttcg gtaccgtcg tgacgttgc 30541 gcctacatcc agaagctcg ggaagaaaac cggaggccgg ctcaggcggtt ggcgcgcaag 30601 attgagtcgg agaaccggta tgccgtgac aacgttcagg cgaggctga ggccgagtc 30661 aagttagtca gccttccacc gtaatggcg gttcccgatc cgttgtggt accgcgtca 30721 cagcgacgac gtcgatctcg ccggacatcg agagcacgtg gaagggtcg ttggccggcg 30781 agagcggcat ccacgcactc gaagacgagt tcgtcacca gttggatcta ggggtcaaga 30841 tcggccgtca cctcaaggat ccgtcgaca ggcacatggg ccgactcgac atgcgacgca 30901 tgcgtacgt ccagcggatg gcaagttgc tggcgccgaca gctatgggat tccggccggca 30961 gcccggaggt cgatccagac cgggtcgccg ttgtgtcg caccggctca ggtggagccg 31021 agaggattgt cgagacgtac gacgtatgc atgcggccgg ccccccggaa gttgtcccg 31081 tggccgttca gatgatcatg cccaaacggtg ccgcggcggt gatcggtcg cagcttgggg 31141 cccgcgcgg ggtgtgacc cgggtgtcg cctgtcgcc gggctcgaa ggcacgcgc 31201 acgcgtggcg tcagatgtc atggcgacg cgcacgtcg cgtctcgcc ggtgtcgaa 31261 gaccatcga ggcgtgccc atcgccgggt tctccatgt gccccatcg acacccgca 31321 acgacgagcc tgagcggcc tccggccgt tgacaaaggaa cggcgacggc ttgtgttcg 31381 gcgaggccgg tgcgtgtatc ctcatcgaga cggaggagca gccaaaggcc cgtggccca

31441 agccgttggc ccgattgctg ggtgccggta tcacctcgga cgccittcat atggtggcgc  
31501 ccgcggccga tgggttcgt gccggtaggg ccatgactcg ctgcgtggag ctggccgggt  
31561 tgtcgccggc ggacatcgac cacgtcaacg cgacacggcac ggacgcgcct atcgccgacg  
31621 ccgcggaggc caacgcacatc cgccgtcgccg gttgtatca ggccgcggtg tacgcggcga  
31681 agtctgcgtt gggccactcg atccgcggc tcgggtcgctt cgagtcgttgc ctcacgggtc  
31741 tgacgctgctg cgcacggcgtc atccgcggcga ccctgaacta cgagacaccc gatcccgaga  
31801 tcgacccgttga cgtcgctgccc ggccgaaccgc gctatggcga ttaccgttac gcagtcaaca  
31861 actcggttgg gttccggcggc cacaatgtgg cgcttgcctt cgggcgttac tgaaggacacga  
31921 catcgccgggtt cgcgaggccc gaggtggggg tccccccgt tgcgggggccc agtccggaccg  
31981 atatgaaagg aacgttcgtca agaccaatga cggagctgtt taccggggaaa gcccgttccct  
32041 acgttgtcgat caccggcattt gccatgacga ccgcgtcgcc gaccgacgcg gagactacgt  
32101 ggaagttgtt gctggaccgc caaagcggga tccgtacgtt ccatgaccca ttgtcgagg  
32161 agttcgacccgtt gccagttcgatccggcgtt atctgttgc ggaattcgac caccagctga  
32221 cgcggatcgatccggcgtt atgggatacc tgcagcggat gtccaccgtt ctgagccggc  
32281 gcctgtggaa aatgcggcgtt caccggagg tggacaccaa tgcattgtt ggttccatcg  
32341 gcaccggcctt ggggtcgccg cggaaactgg ttttcgtt ccatgatatcg cgcgtcgccg  
32401 gaatgaaaggc ggtctcgccg ctgaccgtgc agaagtatcat gccaacggg gccgcccgg  
32461 cggtcgggtt ggaacggcac gccaaggccc ggggtatgc gccggatatcg cgtgtcgcat  
32521 ccggcgccgatccatcgcc cgtcggtggc agcagattgtt gctggagag gccgatggccg  
32581 ccatcgccgatccatcgcc cggcggtggag accaggatcg aagcggttgc ccatcgccggg ttgcgtcaga  
32641 tgcgtatcgatgtccacc aacaacgacg accccgcgg tgcgtatcgcc ccatcgac  
32701 gggaccgcga cggcttgcgtt ttcggcgagg gccggccctt tctgttgc gggaccgg  
32761 agcacgcctt ggcacgtggc gccaacatcc tggccggat catggccgc agcatcac  
32821 ccgttgttccatcggtt gccccggacc ccaacggggac acgcgcggg catcgatcg  
32881 cgcggcgat tcaatcggttgc ggcctcgccc cggcgacat cgaccacgtc aatgcgcac  
32941 ccaccggcac ccaggatcgatcc gacctggccg aaggcaggc catcaacaac gccttggccg  
33001 gcaaccgacc ggcgggttac gcccccaatgttgc ccatcggttgc ggcggccgttgc  
33061 ggcgggttac atcgatcttgc acgggttgc cttgcgttgc tcaatcggttgc cccggac  
33121 tgaatctgtt aaacatcgatcc cccgagatcg atttggacgtt ggtggccgggtt gacccgcac  
33181 cggcaattt cggatcgatcc aataact cttcggttgc cggccggccac aacgtggca  
33241 tcgccttcgg acgggtactaa accccagatcg tgcgtatcg gagacatcg atgacaatca  
33301 tggcccccggc ggcgggttgc ggtcgatcg acccccgatcc tccgtgttgc cgggttgc  
33361 acttcttcgtt cgcacggcgtt gttggatgttgc tgcacgttgc tgcgttgc  
33421 cgcggccggg caccgtcaac ggtgtcgatcc ccatcggttgc ctcacccggac ggcacccgtt  
33481 tggccggccgtt cgcacggcgtt ggggggttgc cgcacatcgatcc caacgccttac gacactgttca

33541 tcgaagacca gagtcccatc gtggcatct ggcattcggg tgggccgg ctggctgaag  
 33601 gtgtcgccc gctcacgca gtaggccagg tggcgaagc catgatccgc gctccggc  
 33661 acatcccgca gatctcggtg gtcgtcggtt tcgcccggg cggccgcgc tacggaccgg  
 5 33721 cggtgaccga cgtcgctgc atggcgccgg aaagccgggt gttcgtaacc gggcccgacg  
 33781 tggcgccag cgtcacccggc gaggacgtcg acatggcctc gtcgggtgg ccggagaccc  
 33841 accacaagaa gtcgggggtg tgccacatcg tcgcccacga cgaactcgat gcctacgacc  
 33901 gtggcgccg gtggcggga ttgttcgtcc agcagggca ttgcgtcg agcaaggccg  
 33961 aggccgggtga caccgacatc cacgcgtgc tgccggaatc ctgcgtacgt gcctacgacg  
 10 34021 tgcgtccgat cgtgacggcg atcctcgatc cggacacacc gttcgacgag ttccaggcca  
 34081 attggcgcc gtcgtatggt gtcggctgg gtcggctgc gggtcgcacg gtgggtgtac  
 34141 tggcaacaa cccgtacgc ctggcgccgt gcctgaactc cggaaagcgca gagaaggcag  
 34201 cgcgttgcgt ggggtgtgc gacgcgttc ggattccgt ggtgggtgt gtcgtatgtc  
 34261 cgggttatct gcccgggtgc gaccaggagt ggggtggcgt ggtgcggccgt ggcgccaagt  
 15 34321 tgctgcacgc gtcggcgag tgcaccgttc cgccggtcac gctggtcacc cggaaagaccc  
 34381 acggcgccgc atacattgcg atgaactccc ggtcggttgcgacg gttcgccct  
 34441 ggccggacgc cgaggctcg ggtatggcgtt ctaaggccgc cgtcgccatc ctgcacaaga  
 34501 agaagtggc cggcgctccg gaggcacaac gcgaaaggcgat gcacgaccag ttggccggc  
 34561 agcatgagcg catgcccggc ggggtcgaca gtgcgttgcgatcgtgt gtcgacgaga  
 20 34621 agatcgaccc ggccgtatct cgcagcaagc tcaccggggc gtcggcgag gtcggccac  
 34681 ggcgcggccg ccacaagaac atcccgctgt agttctgacc gcgagcagac gcagaatcgc  
 34741 acgcgcgagg tccgcggcgt gcgattctgc gtctgtcg cagttatccc cagcgggtggc  
 34801 tggtaacgc gaggcgctcc tgcgtatcgatc ggacgggtggc taccgacgcg ctaacaattc  
 34861 tcgagaaggc cggcggttc gccaccaccc cggatgtct caccgtatg acccgccaac  
 25 34921 agtcgacgt ccaagtggaaa aacggcgcc tcgttgcgtt tgggtacggg gtctacgccc  
 34981 cacaagagcc ggacgttgcg ggcgcgttgg cggctctcgat tgggtatcg gggggccacg  
 35041 ccgtcgctgt tctgggcacc gcccggcgt tggatggatt cgacacggaa aacaccgtcg  
 35101 ctatccatat gtcgtatccc ggagtaagga tgcggccac ggtcggtctg atggtccacc  
 35161 aacgcgttgcg tggccggctc caacgggtgtt cggatgtct cggacccggc cccgcattgg  
 35221 ctgcgttgcg ggtgcgttgc cggatgttc gcccggggc gtcggccacc ctcgacgcgc  
 30 35281 cactacggtc aatgcgttc gtcgtcgatc aaattggaaa cggcgatgtt gaggcagcgag  
 35341 gcccggagg catgcgtcg ggcgcgttgc ctttaccctt cggccgttgcg cggccggaaat  
 35401 cggccatggc gaggcgaggct cggatgttc tgcgtatcgatc cggatgttgcg tggccggac  
 35461 ttcaatcccc gatacaccgc caccgggtgtt aatgtggcg agtcgacttc gcctggcccg  
 35521 acatgcgttgc cggccgttgc aacggatgtt cggatgttgc cggccgttgcg cggatgtc  
 35581 tgcgttgcgttgc gccaagtcgc aacggatgttgc gtcgttgcgttgc gtcgttgcgttgc  
 35

35641 tcgtcgacga tgcagacgc gaacccggcc gcctggcgcc cgcacatcgcc cgccacacctg  
35701 accgcgcgac tggccggc tgaccgctgg tgagcagacg cagatcgca ctgcggccgg  
35761 cgcagtgcga ctctgcgtct gtcgcgctc aacggctgag gaactcctta gccacggcga  
5 35821 ctacgcgctc gcatcccgt ggcaccagac cgatccgggt ccggcggtcg aggatatcg  
35881 ccacatccag cgcggccctca tgggtcaccg cgtattcgaa ctccggccgg gtcacgtcga  
35941 tgccgtcgcc gacggctcg gtggccgct cacatgtggc ggcggcagcg acgttggccg  
36001 cctcgcccccc gtaccgcgccc accagcgact cggcaatcc ggcggccgat ccggggggccg  
36061 gcccagggtt cgccgggtcgcc cgcattcgcc gcagggtcgcc agtgcggcac ttgcggctc  
10 36121 gcagggtcg cagcgtgatg ggcgcattca gcacatcctc tgccatgttag cggtattccg  
36181 tcagcttgcc gccgaccaca ctgatcacgc cgcacggcga ttcaaaaaca gcgtggtcac  
36241 gcgaaacgtc ggcgggtcgcc cctggacac cagcaccgac ggtgtcgatt agcggccgca  
36301 atcccgata ggcacccgtatg acatccctgg tgccgaccgc cgtcccaat ggcgtgttca  
36361 cctgtatcccg caggaacgtt atctctccg aagacgggtt tggcacatcg ggaatcggc  
15 36421 cgggtcgcc ttgcgtcgcc agcccgagat agatccggcc cagctgtcg ggcacggcga  
36481 acacgaagcg gttcagctca cgggggatcg gaatggtcag cgcggcagtc ggattggcaa  
36541 acgacttcgc gtcgaagacc agatgtgtgc cgcggctggg gcgttagcctc agggacgggt  
36601 cgatctcacc cgcggccacacg cccggcggt tgatgcggc acgcggccgac agcgcgaacg  
36661 actgcccgggt ggcgggtcg gtcaactcca ccgaagtgcc ggtgacattc gacgcggcca  
20 36721 cgtaagttag gatgcggcccg cgggtcgcccg cgcgggtcgcc cgcacggcc atgaccagcc  
36781 gggcggtcgcc gatcaattgc cgcgtcgatcg cgagcagacc accgtcgagg cgcgtccggcc  
36841 gaacgggtggg agcaatctcc accacccgtg acgcggggat tggcgccgt cggggcaacg  
36901 tggccggcccg cgtaccccgat agcaccggca aagcgtcgcc ggccaggaaa cggcaccgca  
36961 ccaacgcccccttgggtgtga cccatcgacg gcaacaacgg gaccagttgc ggcacggcat  
25 37021 gcacggatgtt agggcggttg cgtgtcatca ggattcccgcc ttgcacggcg ctgcggccgg  
37081 cgatgcccac gttggcgctg gccagatagc gcagaccggcc gtgcaccaac ttgcagctcc  
37141 agcggctgggt ggcgaacgac agatcatgct tttccaccaa ggccaccgtc agaccgcggg  
37201 tggcagcatc taaggcaatcg ccaacaccgg taatggccgc gcctatcaccg atgacgtcg  
37261 gtgcggccacc gtcggccagt gcggtcgatcg cggcggagcg acgcggcccg ttgagtgcc  
37321 ccgagtggtt catcagcaca aatatccgtt cagtcgtgg gtaagttcg tggccacggc  
37381 ggcggaaatcg aggtcgatcg cgcacgttc cgcggactgg atggcgact gggcgatcg  
37441 caacaccatg gtcggccatgc gacgagcgac gccggagcgac acactgcccc accgctgcgc  
30 37501 cactgtcagc cggccggccca accccctcgat caggacctgc tggctggcgcc cgaggcgctc  
37561 ggtgtatgtac accctggccca gtcggccatgc cggcggagcgac atgatcgat cgtcaccgg  
37621 caaccgggtcg gccaccggcga caatctgcctt taccacacgt tcccggtcg tcccggtcg  
35 37681 gggcacctcc cgcagcgtt cggcgatatg gtcgggtcgcc atggacgcca tgatcgaccg

37741 ggtgtccggc cagcgacggt atacggtcgg ggggctcacg cccgcgcgccc gggcgatctc  
 37801 ggcaagtgtc acccggtcca cgccgtaatc gacgacgcag ctcgcgcgtg cccgcaggat  
 37861 acgaccaccg gtatccgcgc ggtcattact cattgacagc atgtgtataa ctgtaaacgcg  
 5 37921 tgactcaccg cgaggaactc cttccaccga tgaaatggga cgcgtggggat gatccgcgcg  
 37981 cggccaagcc actttctgtat gggtcggtt cgtgtcgaa gcagggttg ggcctagcgg  
 38041 actcggagca gcccgaactc gacccgcgc aggtgcagct ggcgcgtcc gccctgtcgg  
 38101 gggcagacca (SEQ ID NO: 24)

10 **6.9. X-linked Inhibitor of Apoptosis Protein ("XIAP")**  
 GenBank Accession # U45880:

1 gaaaagggtgg acaagtccata ttcaagag aagatgactt ttaacagttt tgaaggatct  
 61 aaaacttgcata tacatgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga  
 121 ttaaaaactt ttgctaaattt tccaagtggt agtccgtttt cagcatcaac actggcacga  
 15 181 gcaggggttc ttatactgg tgaaggagat accgtgcggc gcttttagtt tcatgcagct  
 241 gtagatagat ggcaatatgg agactcagca gttggaaagac acagggaaat atccccaaat  
 301 tgcagattt tcaacggctt ttatcttgaa aatagtgcac cgcagtcac aaattctgg  
 361 atccagaatg gtcagttacaa agttggaaac tatctggaa gcagagatca tttgcctta  
 421 gacaggccat ctgagacaca tgcagactat cttttgagaa ctggcagggt ttagatata  
 481 tcagacacca tatacccgag gaaccctgcc atgtattgtg aagaagctag attaaagtcc  
 541 tttcagaact ggccagacta tgctcacca accccaagag agtttagcaag tgctggactc  
 601 tactacacag gtattggtga ccaagtgcag tgctttgtt gtggggaaa actgaaaaat  
 661 tggttggccctt tgatcggtc ctggcagaa cacaggcgcac actttccat tttgccttt  
 721 gtttggcc ggaatcttaa tattcgaagt gaatctgtat cttgtgatc tgatagaaat  
 781 ttcccaaattt caacaatct tccaagaaat ccatccatgg cagattatga agcacggatc  
 841 tttacttttgg gacatggat atactcgat aacaaggagc agcttgcaag agctggattt  
 901 tatgttttagt gtgaagggtga taaagttaaatg tgccttcactt gtggaggagg gctaactgtat  
 961 tggaagccca gtgaagaccc ttggaaacaa catgttaat ggtatccagg gtgcaaatat  
 1021 ctgttagaac agaaggacaa agaatatata aacaatattc atttaactca ttcacttgag  
 1081 gaggtctgg taagaactac tgagaaaaca ccatcaactaa etagaagaat tgatgatacc  
 1141 atcttccaaa atcctatggt acaagaagct atacgaatgg ggttcagttt caaggacatt  
 1201 aagaaaataa tggaggaaaa aattcagata tctgggagca actataaatac acttgaggat  
 1261 ctgggtgcag atctgtgaa tgctcagaaa gacagtatgc aagatgagtc aagtcagact  
 1321 tcattacaga aagagattag tactgaagag cagctaaaggc gcctgcaaga ggagaagctt  
 1381 tgcaaaaatct gtatggatag aaatattgtat atcgttttttgc ttcctgtgg acatctgtatc  
 1441 acttgtaaac aatgtgcgtga agcagttgac aagtgtccca tgcgtacac agtattact

1501 ttcaagcaaa aaattttat gtcttaatct aactctatag taggcatgtt atgttgttct  
 1561 tattaccctg attgaatgig ttagtgtgaac tgactttaag taatcaggat tgaattccat  
 1621 tagcatttgc taccaagtag gaaaaaaaaat gtacatggca gtgttttagt tggcaatata  
 5 1681 atctttgaat ttcttgattt ttcagggat tagctgtatt atccattttt ttactgtt  
 1741 ttaatgtaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtt  
 1801 attcatagta tactgattt atttctaagt gtaagtgaat taatcatctg gatttttat  
 1861 tctttcaga taggcttaac aaatggagct ttctgtatataaaatgtggag attagagtt  
 1921 atctccccaa tcacataatt tgggggtgtt gaaaaaggaa taaattgttc catgctgggt  
 10 1981 gaaagataga gattgtttt agagggttgtt tggattctg tccattttct  
 2041 tgtaaaggga taaacacgga cgtgtgcgaa atatgtttgtt aaagtgtattt gccatttttg  
 2101 aaagcgtatt taatgtataga atactatcga gccaacatgt actgacatgg aaagatgtca  
 2161 gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gtatgtctg agccagatca  
 2221 aagtatgtat gttgttaata tgcatagaac gagagattt gaaagatata caccaaactg  
 15 2281 ttaaatgtgg ttctcttcg gggagggggg gattggggga ggggccccag aggggtttta  
 2341 gaggggcctt ttcaatttcg actttttca ttttgtctg ttggattttt ttataagtat  
 2401 gtagaccccg aagggtttta tgggaactaa catagtaac ctaacccccc tgactatcc  
 2461 gtgccttc tagggagctg tttttttcc cacccaccac cttttttttt gaacaaatgc  
 2521 ctgagtgctg gggcacttg (SEQ ID NO: 25)

20 General Target Region:

Internal Ribosome Entry Site (IRES) in 5' untranslated region:

5'AGCUCCUAUAACAAAAGUCUGUUGCUGUUUCACAUUUUGGAUU  
 UCCUAUAUAUGUUCUCUUUUAGAAAAGGUGGACAAGUCCUAUUU  
 UCAAGAGAAG3' (SEQ ID NO: 26)

Initial Specific Target Motif:

RNP core binding site within XIAP IRES

5'GGAUUUCCUAUAUAUGUUCUCUUUU3' (SEQ ID NO: 27)

30 **6.10. Survivin**

GenBank Accession # NM\_001168:

1 ccggccagatt tgaatcgccg gaccgggtgg cagaggtggc ggcggccggca tgggtggccc  
 61 gacgttgcggc cctgcctggc agccctttctt caaggaccac cgcatctca cattcaagaa  
 121 ctggcccttc ttggagggtgtt ggcctgcac cccggagcgg atggccgagg ctggcttcat  
 35 181 ccactgcccc actgagaacg agccagactt ggcccaactgt ttcttctgct tcaaggagct

241 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccgggtg  
 301 cgcttcctt tctgtcaaga agcagttga agaattaacc cttggtaat ttttgaact  
 361 ggacagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agaaagaatt  
 5 421 tgaggaaact gcgaagaaag tgcgccgtgc catcgacag ctggctgcca tggattgagg  
 481 cctctggccg gagctgcctg gtcccgagt ggctgcacca cttccagggt ttattccctg  
 541 gtgccaccag ccttcctgtg ggcccttag caatgtctt ggaaaggaga tcaacattt  
 601 caaatttagat gtttcaactg tgctccgtt ttgtcttga agtggcacca gagggtcttc  
 661 tgcctgtgca gcgggtgctg ctggtaacag tggctgcttc tctctctc tctcttttt  
 10 721 gggggctat tttgtgtt ttgattcccg ggcttaccag gtgagaagtg agggaggaag  
 781 aaggcagtgt ccctttgtc agagctgaca gctttgtcg cgtggcaga gccttccaca  
 841 gtgaatgtgt ctggacctca ttttgtttag gctgtcacag tcttgagtgt ggacttggca  
 901 ggtgcctgtt gaatctgagc tgcagggtcc ttatctgtca caccttgcc tcctcagagg  
 961 acatgtttt ttttgtgtt tttttttt ggttagatgca tgacttgtt  
 1021 gtgatgagag aatggagaca gagtccctgg ctccctact gtttaacaac atggctttct  
 1081 tattttgtttaa gaattgttaa ttacagaat agcacaaact acaattaaaaa ctaagcacaa  
 1141 agccattcta agtcattggg gaaacgggtt gaacttcagg tggatgagga gacagaatag  
 1201 agttagatgaa agcgtctggc agataactct ttggccactg ctgtgtgatt agacaggccc  
 1261 agttagccgc gggcacatg ctggccgctc ctccctcaga aaaaggcagt ggcctaaatc  
 20 1321 ctttttaat gacttggctc gatgtgtgg gggactggct gggctgctc aggccgtgt  
 1381 tctgtcagcc caacccac atctgtcact ttccacac gggggagaga cgcagtcgc  
 1441 ccaggtcccc gcttccctt gaggcagcag ctcccgagg gctgaagtct ggcgttaagat  
 1501 gatggatttg attgccttc ctccctgtca tagagctgca gggtgattt ttacagctt  
 1561 gctggaaacc tctggaggc atctcggctg ttccctgagaa ataaaaagcc tgtcatttc (SEQ ID NO: 28)

25

**7. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA  
BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS**

The results presented in this Example indicate that gel mobility shift assays can be used to detect the binding of small molecules, such as the Tat peptide and 30 gentamicin, to their respective target RNAs.

**7.1. Materials and Methods**

**7.1.1. Buffers**

35 Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM MgCl<sub>2</sub>. Tris-borate-EDTA (TBE) buffer is

composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM MgCl<sub>2</sub>.

5

### 7.1.1. Gel retardation analysis

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of either a 5' fluorescein labeled oligonucleotide corresponding to the 16S rRNA A site (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 29); 10 Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448) or a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 15 96:12997-13002) was 3' labeled with 5'-<sup>32</sup>P cytidine 3', 5'-bis(phosphate) (NEN) and T4 RNA ligase (NEBiolabs) in 10% DMSO as per manufacturer's instructions. The labeled oligonucleotides were purified using G-25 Sephadex columns (Boehringer Mannheim). For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was utilized with TK buffer containing 0.5mM MgCl<sub>2</sub> and a 12-mer 20 Tat peptide (YGRKKRRQRRP (SEQ ID NO: 31); single letter amino acid code). For 16S rRNA-gentamicin reactions, the method of Huq *et al.* was used with TKM buffer. In 20 μl reaction volumes 50 pmoles of <sup>32</sup>P cytidine-labeled oligonucleotide and either gentamicin sulfate (Sigma) or the short Tat peptide (Tat<sub>47-58</sub>) in TK or TKM buffer were heated at 90°C for 2 minutes and allow to cool to room temperature (approximately 24°C) 25 over 2 hours. Then 10 μl of 30% glycerol was added to each reaction tube and the entire sample was loaded onto a TBE non-denaturing polyacrylamide gel and electrophoresed at 1200-1600 volt-hours at 4°C. The gel was exposed to an intensifying screen and radioactivity was quantitated using a Typhoon phosphorimager (Molecular Dynamics).

30

### 7.2. Background

One method used to demonstrate small molecule interactions with natural occurring RNA structures such as ribosomes is by a method called chemical footprinting or toe printing (Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448). Here the use of gel 35 mobility shift assays to monitor RNA-small molecule interactions are described. This approach allows for rapid visualization of small molecule-RNA interactions based on the

difference between mobility of RNA alone versus RNA in a complex with a small molecule. To validate this approach, an RNA oligonucleotide corresponding to the well-characterized gentamicin binding site on the 16S rRNA (Moazed & Noller, 1987, *Nature* 327:389-394) and the equally well-characterized HIV-1 TAT protein binding site on the HIV-1 TAR element (Huq *et al.*, 1999, *Nucleic Acids Res.* 27: 1084-1093) were chosen. The purpose of these experiments is to lay the groundwork for the use of chromatographic techniques in a high throughput fashion, such as microcapillary electrophoresis, for drug discovery.

10

### 7.3. Results

A gel retardation assay was performed using the Tat<sub>47-58</sub> peptide and the TAR RNA oligonucleotide. As shown in Figure 1, in the presence of the Tat peptide, a clear shift is visible when the products are separated on a 12% non-denaturing polyacrylamide gel. In the reaction that lacks peptide, only the free RNA is visible. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Huq *et al.*, 1999, *Nucleic Acids Res.* 27: 1084-1093).

Based on the results of Figure 1, it was hypothesized that RNA interactions with small organic molecules could also be visualized using this method. As shown in Figure 2, the addition of varying concentrations of gentamicin to an RNA oligonucleotide corresponding to the 16S rRNA A site produces a mobility shift. These results demonstrate that the binding of the small molecule gentamicin to an RNA oligonucleotide having a defined structure in solution can be monitored using this approach. In addition, as shown in Figure 2, a concentration as low as 10ng/ml gentamicin produces the mobility shift.

To determine whether lower concentrations of gentamicin would be sufficient to produce a gel shift, similar experiment was performed, as shown in Figure 2, except that the concentrations of gentamicin ranged from 100 ng/ml to 10 pg/ml. As shown in Figure 3, gel mobility shifts are produced when the gentamicin concentration is as low as 10 pg/ml. Further, the results shown in Figure 3 demonstrate that the shift is specific to the 16S rRNA oligonucleotide as the use of an unrelated oligonucleotide, corresponding to the HIV TAR RNA element, does not result in a gel mobility shift when incubated with 10  $\mu$ g/ml gentamicin. In addition, if a concentration as low as 10 pg/ml gentamicin produces a gel mobility shift then it should be possible to detect changes to RNA structural motifs when small amounts of compound from a library of diverse compounds is screened in this fashion.

5                   Further analysis of the gentamicin-RNA interaction indicates that the interaction is Mg- and temperature dependent. As shown in Figure 4, when MgCl<sub>2</sub> is not present (TK buffer), 1mg/ml of gentamicin must be added to the reaction to produce a gel shift.

10                  Similarly, the temperature of the reaction when gentamicin is added is also important. When gentamicin is present in the reaction during the entire denaturation/renaturation cycle, that is, when gentamicin is added at 90°C or 85°C, a gel shift is visualized (data not shown). In contrast, when gentamicin is added after the renaturation step has proceeded to 75°C, a mobility shift is not produced. These results are 15 consistent with the notion that gentamicin may recognize and interact with an RNA structure formed early in the renaturation process.

20                  8. EXAMPLE:     **IDENTIFICATION OF A DYE-LABELED TARGET RNA  
15                  BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS  
BY CAPILLARY ELECTROPHORESIS**

---

The results presented in this Example indicate that interactions between a peptide and its target RNA, such as the Tat peptide and TAR RNA, can be monitored by gel retardation assays in an automated capillary electrophoresis system.

25                  8.1. **Materials and Methods**

30                  8.1.1. **Buffers**

Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM MgCl<sub>2</sub>. Tris-borate-EDTA (TBE) buffer is 25 composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM MgCl<sub>2</sub>.

35                  8.1.1. **Gel Retardation Analysis Using Capillary Electrophoresis**

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUUCGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was used. For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was

utilized with TK buffer containing 0.5mM MgCl<sub>2</sub> and a 12-mer Tat peptide (YGRKKRRQRRP (SEQ ID NO: 31); single letter amino acid code). In 20  $\mu$ l reaction volumes 50 pmoles of labeled oligonucleotide and the short Tat peptide (Tat<sub>47-58</sub>) in TK or 5 TKM buffer were heated at 90°C for 2 minutes and allow to cool to room temperature (approximately 24°C) over 2 hours. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania).

### 8.2. Results

10 As presented in the previous Example in Section 7, interactions between a peptide and RNA can be monitored by gel retardation assays. It was hypothesized that interactions between a peptide and RNA could be monitored by gel retardation assays by an automated capillary electrophoresis system. To test this hypothesis, a gel retardation assay by an automated capillary electrophoresis system was performed using the Tat<sub>47-58</sub> peptide 15 and the TAR RNA oligonucleotide. As shown in Figure 5 using the capillary electrophoresis system, in the presence of the Tat peptide, a clear shift is visible upon the addition of increasing concentrations of Tat peptide. In the reaction that lacks peptide, only a peak corresponding to the free RNA is observed. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 20 94:3548-3553; Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093).

25 The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

30

35

The invention can be illustrated by the following embodiments enumerated in the numbered paragraphs that follow:

- 5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA: test compound complex is formed; (b) separating the detectably labeled target RNA: test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds; and (c) determining a structure of the test compound bound to the RNA in the RNA: test compound complex.
- 10 2. The method of paragraph 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or adenylate uridylate-rich element.
- 15 3. The method of paragraph 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- $\alpha$ "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.
- 20 4. The method of paragraph 1 in which the detectably labeled RNA is labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.
- 25 5. The method of paragraph 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries, including but not limited to, libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

6. The method of paragraph 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions.

5

7. The method of paragraph 6 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.

10

8. The method of paragraph 6 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. In another embodiment, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl<sub>2</sub>. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl<sub>2</sub>. In another embodiment, the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

15

9. Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids in the method of paragraph 1. In a preferred embodiment, electrophoresis is used for separation of the complexed and non-complexed target nucleic acids. In a preferred embodiment, the electrophoresis is capillary electrophoresis. In other embodiments, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation are used for the separation of the complexed and non-complexed target nucleic acids.

20

10. The structure of the test compound of the RNA:test compound complex of paragraph 1 is determined, in part, by the type of library of test compounds. In a preferred embodiment wherein the combinatorial libraries are small organic molecule libraries, mass spectroscopy, NMR, or vibration spectroscopy are used to determine the structure of the test compounds.

25

30

## WHAT IS CLAIMED IS:

1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:

5 (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA: test compound complex is formed;

10 (b) separating the detectably labeled target RNA: test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds by capillary gel electrophoresis; and

15 (c) determining a structure of the test compound bound to the RNA in the RNA: test compound complex by mass spectroscopy.

20

25

30

35

**AMENDED CLAIMS**

[received by the International Bureau on 17 September 2002 (17.09.02);  
Claims 1 to 10 replaced by new claims 1 to 19. (3 sheets)]

5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:

- (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA: test compound complex is formed;
- 10 (b) separating the detectably labeled target RNA: test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds; and
- 15 (c) determining a structure of the test compound bound to the RNA in the RNA: test compound complex.

2. The method of claim 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or  
20 adenylate uridylate-rich element.

3. The method of claim 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- $\alpha$ "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6  
25 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.

4. The method of claim 1 in which the detectably labeled RNA is  
30 labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.

5. The method of claim 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as  
35 hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal

peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; or small organic molecule libraries.

5

6. The method of claim 5 in which the small organic molecule libraries are libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

10

7. The method of claim 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution wherein the aqueous solution comprises a buffer and a combination of salts.

15

8. The method of claim 7 wherein the aqueous solution approximates or mimics physiologic conditions.

20

9. The method of claim 7 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.

10. The method of claim 7 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant.

25

11. The method of claim 10 in which the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl<sub>2</sub>.

30

13. The method of claim 11 wherein the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl<sub>2</sub>.

14. The method of claim 10 wherein the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

35

15. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and 5 test compounds is by electrophoresis.

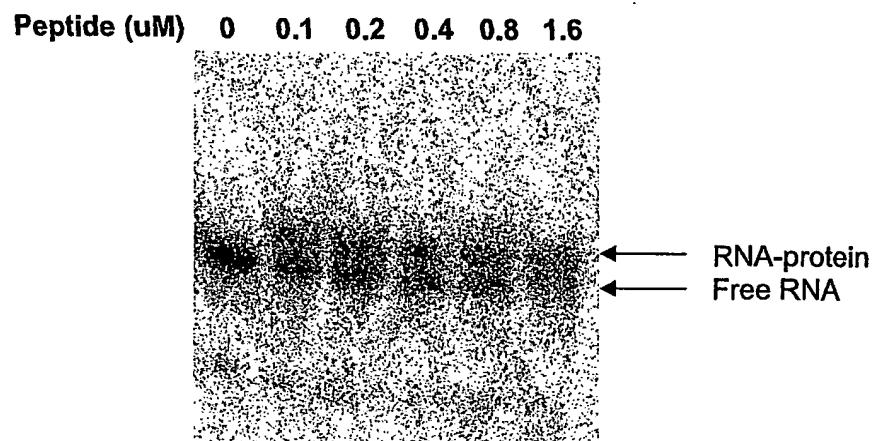
16. The method of claim 15 in which the electrophoresis is capillary electrophoresis.

10 17. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and test compounds is by fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, or 15 nanoparticle aggregation.

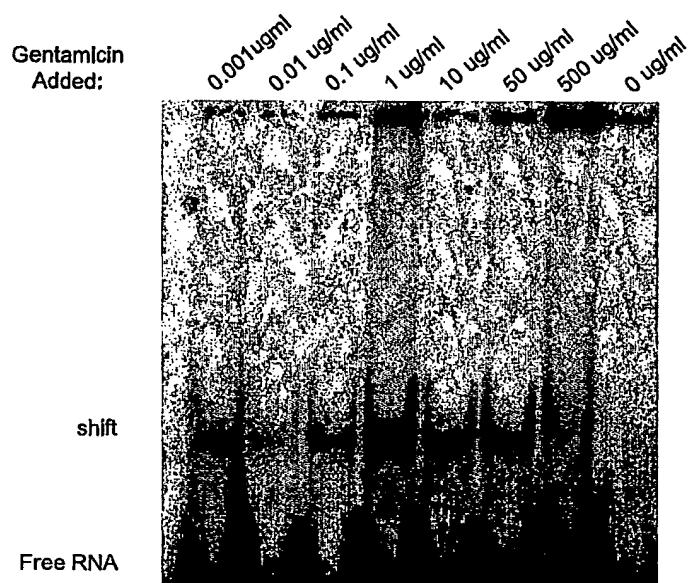
18. The method of claim 1 in which the library of test compounds are small organic molecule libraries.

20 19. The method of claim 18 in which the structure of the test compound is determined by mass spectroscopy, NMR, or vibration spectroscopy.

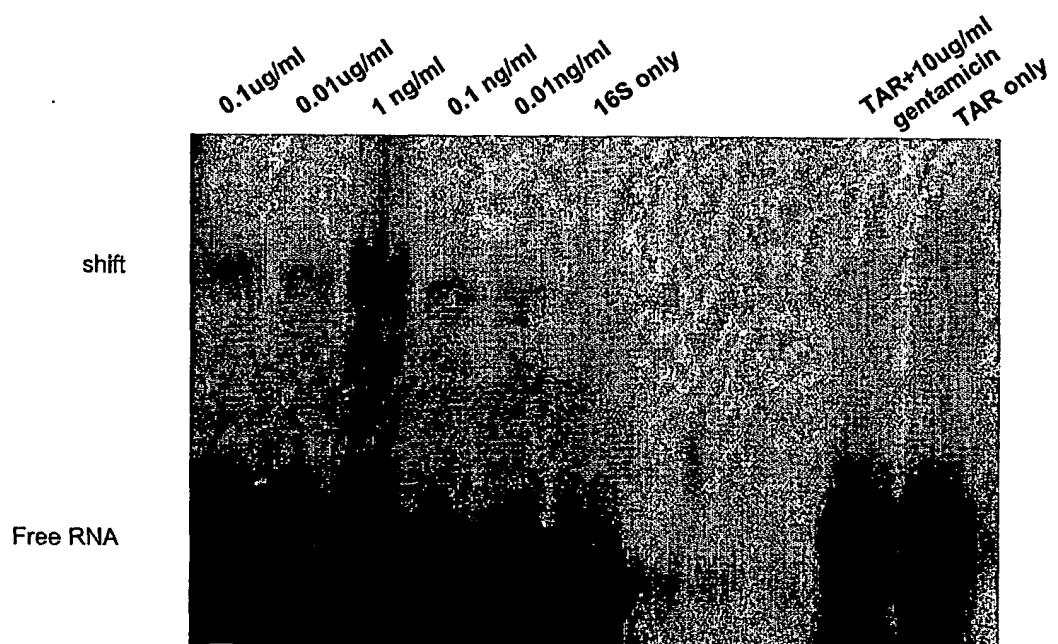
**Figure 1**  
**Sheet 1/5**  
**Attorney Docket No. 10589-007**



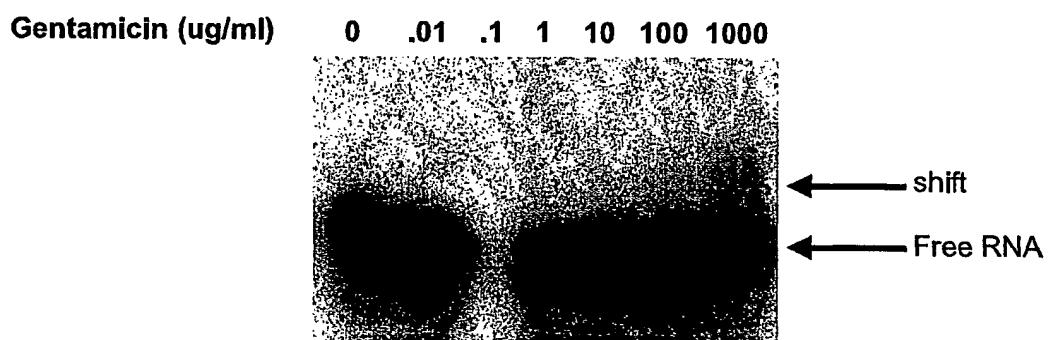
**Figure 2**  
**Sheet 2/5**  
**Attorney Docket No. 10589-007**



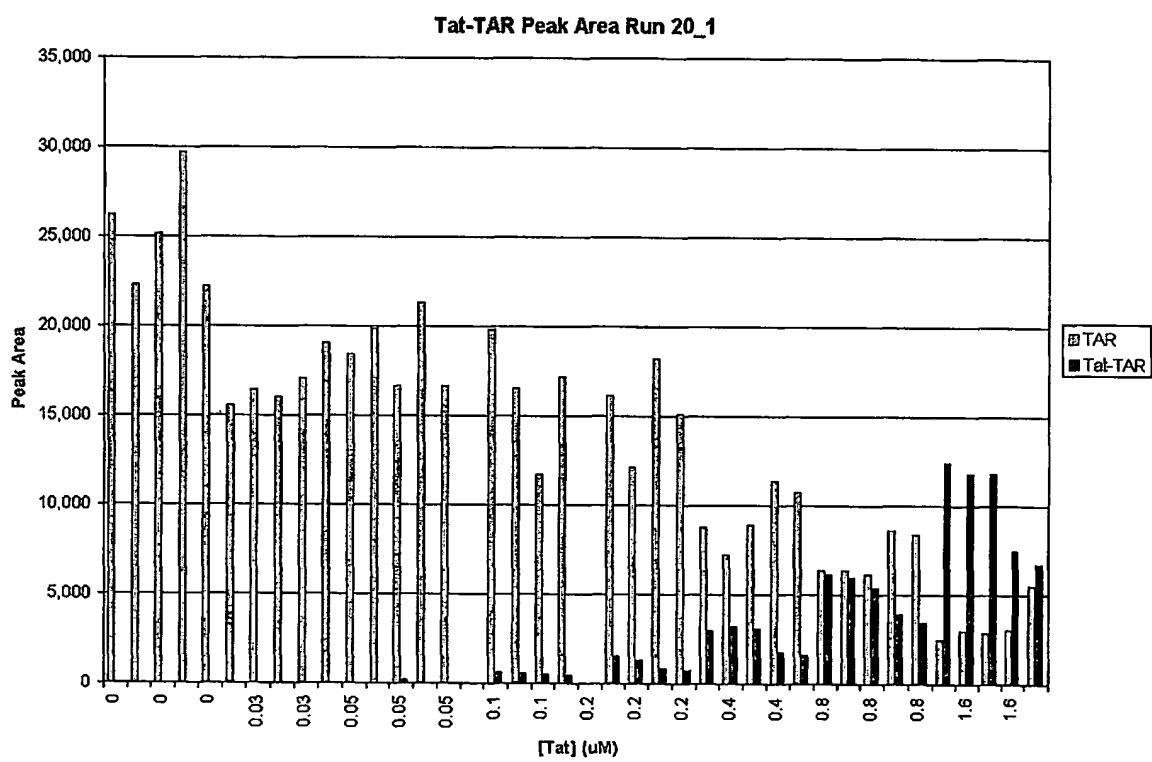
**Figure 3**  
**Sheet 3/5**  
**Attorney Docket No. 10589-007**



**Figure 4**  
**Sheet 4/5**  
**Attorney Docket No. 10589-007**



**Figure 5**  
**Sheet 5/5**  
**Attorney Docket No. 10589-007**



## SEQUENCE LISTING

<110> PTC Therapeutics, Inc.

<120> METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

<130> 10589-007-228

<140> To be assigned

<141> 2002-04-11

<150> 60/282,965

<151> 2001-04-11

<160> 31

<170> PatentIn version 3.0

<210> 1

<211> 21

<212> RNA

<213> Homo sapiens

<400> 1

auuuauuuau uuauuuuauuu a

21

<210> 2

<211> 17

<212> RNA

<213> Homo sapiens

<400> 2

auuuauuuau uuauuuua

17

<210> 3  
<211> 15  
<212> RNA  
<213> Homo sapiens

<400> 3  
wauuuauuuu uuuaw

15

<210> 4  
<211> 13  
<212> RNA  
<213> Homo sapiens

<400> 4  
wwauuuauuu aww

13

<210> 5  
<211> 13  
<212> RNA  
<213> Homo sapiens

<400> 5  
wwwauuuaw www

13

<210> 6  
<211> 1643  
<212> DNA  
<213> Homo sapiens

<400> 6  
gcagaggacc agctaagagg gagagaagca actacagacc cccccctgaaa acaaccctca 60  
gacgccacat cccctgacaa gctgccaggc aggttctctt cctctcacat actgaccac 120  
ggctccaccc tctctccctt ggaaaggaca ccatgagcac tgaaaggatg atccgggacg 180  
tggagctggc cgaggaggcg ctccccaga agacaggggg gccccaggc tccaggcggt 240  
gcttggtcct cagcctttc tccttcctga tcgtggcagg cgccaccacg ctcttctgcc 300

tgctgactt tggagtgatc ggcccccaga gggaaagagtt ccccaggac ctctctctaa	360
tcagccctct ggcccaggca gtcagatcat cttctcgaac cccgagtgc aagcctgtag	420
cccatgttgt agcaaaccct caagctgagg ggcagctcca gtggctgaac cgccgggcca	480
atgcctcctt ggccaatggc gtggagctga gagataacca gctgggtggc ccatcagagg	540
gcctgtacct catctactcc caggtccctct tcaaggccca aggctgcccc tccacccatg	600
tgctcctcac ccacaccatc agccgcacatcg ccgtctccta ccagaccaag gtcaacctcc	660
tctctgccat caagagcccc tgccagaggg agacccaga gggggctgag gccaagccct	720
ggtatgagcc catctatctg ggaggggtct tccagctgga gaagggtgac cgactcagcg	780
ctgagatcaa tcggcccgac tatctcgact ttgccgagtc tggcagggc tactttggga	840
tcattgccct gtgaggagga cgaacatcca accttcccaa acgcctcccc tgccccaatc	900
cctttattac cccctccctc agacaccctc aacctcttct ggctaaaaa gagaattggg	960
ggcttagggt cggaaacccaa gcttagaact ttaagcaaca agaccaccac ttgcggaaacct	1020
gggattcagg aatgtgtggc ctgcacagtg aattgtgtgc aaccactaag aattcaaact	1080
ggggcctcca gaactcactg gggcctacag ctttgatccc tgacatctgg aatctggaga	1140
ccagggagcc tttgggtctg gccagaatgc tgcaggactt gagaagacct cacctagaaa	1200
ttgacacaag tggacacttag gccttcctct ctccagatgt ttccagactt ctttgagaca	1260
cggagcccaag ccctccccat ggagccagct ccctctattt atgtttgcac ttgtgattat	1320
ttattattta ttattatttt atttattttc agatgaatgt atttattttgg gagaccgggg	1380
tatcctgggg gacccaatgt aggagctgcc ttggctcaga catgtttcc gtggaaacgg	1440
agctgaacaa taggctgttc ccatgttagcc ccctggcctc tgccttcattt tttgattatg	1500
ttttttaaaa tatttatctg attaagttgt ctaaacaatg ctgatttggt gaccaactgt	1560
cactcattgc tgagcctctg ctccccaggg gagttgtgtc tgtaatcgcc ctactattca	1620
gtggcgagaa ataaagtttgc ctt	1643

&lt;210&gt; 7

&lt;211&gt; 756

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<400> 7	
gctggaggat gtggctgcag agcctgctgc tcttggcac tgtggcctgc agcatctctg	60
cacccggcccg ctcggccagc cccagcacgc agccctggga gcatgtgaat gccatccagg	120
aggcccccggc tctcctgaac ctgagtagag acactgctgc tgagatgaat gaaacagtag	180

aagtcatctc agaaatgttt gacctccagg agccgacctg cctacagacc cgcctggagc	240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg	300
ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtgca acccagacta	360
tcaccttga aagtttcaaa gagaacotga aggacttct gcttgcata cccttgact	420
gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc	480
tctctcatga aacaagagct agaaactcag gatggcatac ttggaggac caaggggtgg	540
gccacagcca tggggaggt ggcctggacc tgccctggc cacactgacc ctgatacagg	600
catggcagaa gaatggaat attttatact gacagaaatc agtaatattt atatatttat	660
atttttaaaa tattttatata tttatattt taagttcata ttccatattt attcaagatg	720
ttttaccgta ataattatta taaaaatata gcttct	756

<210> 8

<211> 756

<212> DNA

<213> Homo sapiens

<400> 8	
tctggaggat gtggctgcag agcctgctgc tcttggcac tgtggcctgc agcatctctg	60
cacccgcccc ctcgccccagc cccagcacgc agccctggga gcatgtaat gccatccagg	120
aggcccccgc tctcctgaac ctgagtagag acactgctgc tgagataat gaaacagtag	180
aagtcatctc agaaatgttt gacctccagg agccgacctg cctacagacc cgcctggagc	240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg	300
ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtgca acccagacta	360
tcaccttga aagtttcaaa gagaacotga aggacttct gcttgcata cccttgact	420
gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc	480
tctctcatga aacaagagct agaaactcag gatggcatac ttggaggac caaggggtgg	540
gccacagcca tggggaggt ggcctggacc tgccctggc cacactgacc ctgatacagg	600
catggcagaa gaatggaat attttatact gacagaaatc agtaatattt atatatttat	660
atttttaaaa tattttatata tttatattt taagttcata ttccatattt attcaagatg	720
ttttaccgta ataattatta taaaaatata gcttct	756

<210> 9

<211> 825

<212> DNA

<213> Homo sapiens

<400> 9

atcactctct ttaatcacta ctcacattaa cctcaactcc tgccacaatg tacaggatgc	60
aactcctgtc ttgcattgca ctaattttg cacttgtcac aaacagtgc cctacttcaa	120
gttcgacaaa gaaaacaaag aaaacacagc tacaactgga gcatttactg ctggattac	180
agatgatttt gaatggaatt aataattaca agaatccaa actcaccagg atgctcacat	240
ttaagttta catgcccag aaggccacag aactgaaaca gcttcagtgt ctagaagaag	300
aactcaaacc tctggaggaa gtgctgaatt tagctcaaag caaaaacttt cacttaagac	360
ccagggactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaaacaa	420
cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga	480
ttaccttttgc tcaaaggcatc atctcaacac taacttgata attaagtgt tccacttaa	540
aacatatcag gccttctatt tatttattta aatatttaaa ttttatattt attgttgaat	600
gtatggttgc tacctattgt aactattatt cttaatctta aaactataaa tatggatctt	660
ttatgattct ttttgaagc cctagggct ctaaaatggt ttaccttatt tatccaaaa	720
atatttatta ttatgttcaa tgttaaatat agtatctatg tagattgggt agtaaaaacta	780
tttaataaaat ttgataaaata taaaaaaaaaa aaacaaaaaaaaaa aaaaaa	825

<210> 10

<211> 15

<212> RNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)..(1)

<223> N = A, U, G, OR C

<220>

<221> misc\_feature

<222> (15)..(15)

<223> N = A, U, G, OR C

<400> 10  
nauuuauuuua uuuuan

15

<210> 11  
<211> 1125  
<212> DNA  
<213> Homo sapiens

<400> 11  
ttctgcctc gagcccaccc ggaacgaaag agaagctcta tctgcctcc aggagcccag 60  
ctatgaactc cttctccaca agcgccctcg gtccagttgc cttctccctg gggctgctcc 120  
tggtgttgc tgctgcctc cctgccccag taccccccagg agaagattcc aaagatgttag 180  
ccgccccaca cagacagcca ctcacctt cagaacgaat tgacaaacaa attcggtaca 240  
tcctcgacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa 300  
gcagcaaaga ggcactggca gaaaacaacc tgaaccttcc aaagatggct gaaaaagatg 360  
gatgcttcca atctggattc aatgaggaga cttgcctggt gaaaatcatc actggcttt 420  
tggagttga ggtataccta gagtacctcc agaacagatt tgagagtagt gaggaacaag 480  
ccagagctgt gcagatgagt acaaaagtcc tgatccagtt cctgcagaaa aaggcaaaga 540  
atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc 600  
aggcacagaa ccagtggctg caggacatga caactcatct cattctgcgc agctttaagg 660  
agttcctgca gtccagcctg agggctttc ggcaaatgta gcatggcac ctcagattgt 720  
tgggttaat gggcattcct tcttctggtc agaaacctgt ccactggca cagaacttat 780  
gttgttctct atggagaact aaaagtatga gcgttaggac actatttaa ttatTTTaa 840  
tttattaata tttaaatatg tgaagctgag ttaatTTTatg taagtcatat ttatTTTtt 900  
aagaagtacc acttggaaaca ttttatgtat tagtttggaa ataataatgg aaagtggcta 960  
tgcagttga atatcctttt tttcagagcc agatcatttc ttggaaagtg taggcttacc 1020  
tcaaataaat ggctaactta tacatattt taaagaata tttatattgt atttatataa 1080  
tgtataaatg gttttatac caataaatgg cattttaaaa aattc 1125

<210> 12  
<211> 3166  
<212> DNA  
<213> Homo sapiens

<400> 12  
aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagac gcgcggcgt 60  
gcgagcagcg aaagcgacag gggcaaagtg agtacactgc ttttgggggt gaccgccga 120  
gcgcggcgtg agccctcccc cttggatcc cgcagctgac cagtcgcgt gacggacaga 180  
cagacagaca ccgccccca gccccagttac caccctctcc ccggccggcg gcggacagt 240  
gacgcggcgg cgagccgcgg gcaggggccc gagcccgccc ccggaggcgg ggtggaggg 300  
gtcggagctc gcggcgtcgc actgaaactt ttcttccaaac ttctgggtg ttctcgcttc 360  
ggaggagccg tggccgcgc gggggaaagcc gagccgagcg gagccgcgag aagtgttagc 420  
tcggccggg aggagccgca gcgggaggag ggggaggagg aagaagagaa ggaagaggag 480  
agggggccgc agtggcgcact cggcgctcgg aagccggct catggacggg tgaggcggcg 540  
gtgtgcgcag acagtgcgtcc agcgcgcgcg ctccccagcc ctggccgcgc ctggggccgg 600  
gaggaagagt agctcgccga ggcgcgagg agagcgggcc gccccacagc ccgagccgga 660  
gagggacgcg agccgcgcgc cccggtcggg cctccgaaac catgaacttt ctgtgttctt 720  
gggtgcattt gagcatttgc ttgtgtgtctt acctccacca tgccaaatgg tcccaaggctg 780  
cacccatggc agaaggagga gggcagaatc atcacgaatg ggtaaatgtc atggatgtct 840  
atcagcgcag ctactgcattt ccaatcgaga ccctgggtga catcttccag gagtaccctg 900  
atgagatcga gtacatcttc aagccatctt gtgtggccct gatgcgtatgc gggggctgct 960  
ccaatgacga gggcattggag tttgtgtccca ctgaggagtc caacatcacc atgcagat 1020  
tgcggatcaa acctcaccaa ggcgcgcaca taggagat gagcttccta cagcacaaca 1080  
aatgtaatg cagaccaaag aaagatagag caagacaaga aaatccctgt gggcattgt 1140  
cagagcggag aaagcatttg tttgtacaatg atccgcgcac gtgtaaatgt tcctgcaaaa 1200  
acacacactc gctttgcgaag ggcggcgcgc ttgtttttt cgaacgtact tgcagatgt 1260  
acaagccgag gcccggcggcc gggcaggagg aaggagcctc cctcagggtt tcgggaaacca 1320  
gatctctctc cagggaaagac tgatacagaa cgcgcatac agaaaccacg ctggccgcac 1380  
cacaccatca ccatcgacacg aacagtccctt aatccagaaaa cctgaaatga aggaagagga 1440  
gactctgcgc agagcacttt gggccggag ggcgcgcac cggcggaaacg attccgggc 1500  
gggtgacccca gcacggtccc tcttggattt ggattcgcca ttttattttt cttgtgtcta 1560  
aatcaccgag cccggaaatg tagagatttt tatttctggg attcctgtatg acacacccac 1620  
ccacatacat acatttatat atatatatat tatatatata taaaaataaa tatctctatt 1680  
ttatatatat aaaatataaa tattttttt ttaaattaac agtgcataatg ttattgggt 1740  
cttcactgga tgtatggac tgctgtggac ttgagttggg agggaaatgt tcccactcag 1800

atcctgacag ggaagaggag gagatgagag actctggcat gatcttttt ttgtcccaact	1860
tggtgccgc agggtcctct cccctgccc agaatgtgca aggccagggc atggggcaa	1920
atatgaccca gtttggaa caccgacaaa cccagccctg gcgctgagcc tctctacccc	1980
aggtcagacg gacagaaaga caaatcacag gttccggat gaggacaccg gctctgacca	2040
ggagtttggg gagcttcagg acattgtgt gctttggga ttccctccac atgctgcacg	2100
cgcacatctgc ccccagggc actgcctgga agattcagga gcctgggogg cttcgctta	2160
ctctcacctg cttctgagtt gcccaggagg ccactggcag atgtcccgac gaagagaaga	2220
gacacattgt tggaagaagc agcccatgac agcgcccctt cctggactc gccctcatcc	2280
tcttcctgct ccccttcctg gggtcagcc taaaaggacc tatgtcctca caccattgaa	2340
accactagtt ctgtccccc aggaaacctg gttgtgtgt tttgagtgt tgacccctt	2400
ccatccctg gtccttcctt tccctcccg aggacacagag agacagggca ggatccacgt	2460
gcccatgtg gaggcagaga aaagagaaaag ttttttat acggtaactt ttaatatcc	2520
cttttaatt agaaattaga acagtttaatt taattaaaga gtagggttt ttttcagtt	2580
tcttggtaa tatttaattt caactattt tgagatgtat ctttgtct ctctgtct	2640
cattttgtta cgggttttg tatataaaat tcatgtttcc aatctcttc tccctgatcg	2700
gtgacagtca ctatgttatac ttgaacagat atttaattt gctaacaactc agctctgccc	2760
tcccccgttcc cctggctccc cagcacacat tccttgaaa gagggtttca atatacatct	2820
acatactata tatatattgg gcaacttgtt tttgtgtgt tataatata tataatgttt	2880
tgtatatatg tgcgttgc gaaataaaca tcgttattct gtttttata tggtaaaacc	2940
aaacaagaaa aaatagagaa ttctacatac taaatcttc tccttttta attttat	3000
ttgttatcat ttatttattt gtcgttactgt ttatccgtaa taattgtgg gaaaagat	3060
taacatcactt tctttgttcc tagtgcagtt ttgcagata ttccgtagta catatttt	3120
tttaaacaac gacaaagaaa tacagatata tcttaaaaaa aaaaaa	3166

&lt;210&gt; 13

&lt;211&gt; 249

&lt;212&gt; RNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 13

ccgggcucau ggacggugaa ggcggcgug ugcccggaca gugcuccagc ggcgcgcuc	60
cccagccug gcccggccuc gggccggag gaagaguagc ucgccggagc gcccggaga	120
gcgggcccgc ccacagcccg agccggagag ggacgcgagc cgcgcgcggccu	180

ccgaaaccau	gaacuuucug	cugucuuggg	ugcauuggag	ccuugccuug	cugcucuacc	240
uccaccaug						249

<210> 14

<211> 9181

<212> DNA

<213> Homo sapiens

<400> 14						
ggtcctctcg	gttagaccag	atctgagcct	gggagcttc	tggctaacta	ggaaacccac	60
tgcttaagcc	tcaataaaagc	ttgccttgag	tgcttcaagt	agtgtgtgcc	cgtctgttgt	120
gtgactctgg	taactagaga	tccctcagac	ccttttagtc	agtgtggaaa	atctctagca	180
gtggcgcccg	aacagggacc	tgaaagcgaa	agggaaacca	gaggagctct	ctcgacgcag	240
gactcggctt	gctgaagcgc	gcacggcaag	aggcgagggg	cggcactgg	tgagtacgcc	300
aaaaattttg	actagcggag	gctagaagga	gagagatggg	tgcgagagcg	tcagtattaa	360
gcgggggaga	attagatcga	tggaaaaaaa	ttcggttaag	gccaggggga	aagaaaaat	420
ataaattaaa	acatatagtt	tggcaagca	gggagctaga	acgattcgca	gttaatcctg	480
gcctgttaga	aacatcagaa	ggctgttagac	aaatactggg	acagctacaa	ccatcccttc	540
agacaggatc	agaagaactt	agatcattat	ataatacagt	accaaccctc	tattgtgtgc	600
atcaaaggat	agagataaaa	gacaccaagg	aagctttaga	caagatagag	gaagagcaaa	660
acaaaagtaa	gaaaaaaagca	cagcaagcag	cagctgacac	aggacacagc	aatcaggtca	720
gccaaaatta	ccctatactg	cagaacatcc	agggcaaat	ggtacatcag	gccatatcac	780
ctagaacttt	aaatgcattgg	gtaaaagtag	tagaagagaa	ggcttcagc	ccagaagtga	840
tacccatgtt	ttcagcatta	tcagaaggag	ccaccccaca	agatttaaac	accatgctaa	900
acacagtgg	gggacatcaa	gcagccatgc	aatgttaaa	agagaccatc	aatgaggaag	960
ctgcagaatg	ggatagagtg	catccagtgc	atgcagggcc	tattgcacca	ggccagatga	1020
gagaaccaag	gggaagtgac	atagcaggaa	ctactagtac	cttcagggaa	caaataggat	1080
ggatgacaaa	taatccacct	atcccagttag	gagaaattta	taaaagatgg	ataatcctgg	1140
gattaaataa	aatagtaaga	atgtatagcc	ctaccagcat	tctggacata	agacaaggac	1200
caaaggaacc	ctttagagac	tatgttagacc	ggttctataa	aactctaaga	gccgagcaag	1260
cttcacagga	ggtaaaaaat	tggatgacag	aaaccttggtt	ggtccaaaat	gcgaacccag	1320
attgtaagac	tattttaaaa	gcattgggac	cagcggctac	actagaagaa	atgatgacag	1380
catgtcaggg	agttaggagga	cccgccata	aggcaagagt	tttggctgaa	gcaatgagcc	1440

aagtaacaaa	ttaagctacc	ataatgatgc	agagaggcaa	ttttaggaac	caaagaaaga	1500
ttgttaagtg	tttcaattgt	ggcaaagaag	ggcacacagc	cagaaattgc	agggccccta	1560
ggaaaaaggg	ctgttggaaa	tgtggaaagg	aaggacacca	aatgaaagat	tgtactgaga	1620
gacaggctaa	tttttaggg	aagatctggc	cttcctacaa	gggaaggcca	gggaattttc	1680
ttcagagcag	accagagcca	acagccccac	cagaagagag	cttcaggctc	ggggtagaga	1740
caacaactcc	ccctcagaag	caggagccga	tagacaagga	actgtatcct	ttaacttccc	1800
tcaggtcact	cttggcaac	gaccctcggt	cacaataaag	atagggggc	aactaaagga	1860
agctctatta	gatacaggag	cagatgatac	agtattagaa	gaaatgagtt	tgccaggaag	1920
atggaaacca	aaaatgatag	gggaaattgg	aggtttatc	aaagtaagac	agtatgatca	1980
gatactcata	gaaatctgtg	gacataaagc	tataggtaca	gtattagtag	gacctacacc	2040
tgtcaacata	atttggaaagaa	atctgttgc	tcagattgg	tgcactttaa	atttcccat	2100
tagccctatt	gagactgtac	cagtaaaatt	aaagccagga	atggatggcc	caaaagttaa	2160
acaatggcca	ttgacagaag	aaaaaataaa	agcattagta	gaaatttgt	cagagatgga	2220
aaaggaaggg	aaaatttcaa	aaattgggcc	tgaaaatcca	tacaatactc	cagtattgc	2280
cataaagaaa	aaagacagta	ctaaatggag	aaaatttagta	gatttcagag	aacttaataa	2340
gagaactcaa	gacttctggg	aatgtcaatt	aggaatacca	catcccgag	ggttaaaaaa	2400
aaaaaaatca	gtaacagtac	tggatgtgg	tgatgcata	tttcagttc	ccttagatga	2460
agacttcagg	aagtatactg	catttaccat	acctagtata	aacaatgaga	caccaggat	2520
tagatatcag	tacaatgtgc	ttccacaggg	atggaaagga	tcaccagcaa	tattccaaag	2580
tagcatgaca	aaaatcttag	agccttttag	aaaacaaaat	ccagacatag	ttatctatca	2640
atacatggat	gatttgtatg	taggatctga	cttagaaata	ggcagcata	gaacaaaaat	2700
agaggagctg	agacaacatc	tgttgaggtg	ggacttacc	acaccagaca	aaaaacatca	2760
gaaagaacct	ccattcctt	ggatgggta	tgaactccat	cctgataaaat	ggacagtaca	2820
gcctatagtg	ctgccagaaa	aagacagctg	gactgtcaat	gacatacaga	agtttagtgg	2880
gaaattgaat	tggcaagtc	agatttaccc	aggattaaa	gtaaggcaat	tatgtaaact	2940
cottagagga	accaaagcac	taacagaagt	aataccacta	acagaagaag	cagagctaga	3000
actggcagaa	aacagagaga	ttctaaaaga	accagtacat	ggagtgtatt	atgaccatc	3060
aaaagactta	atagcagaaa	tacagaagca	ggggcaaggc	caatggacat	atcaaattta	3120
tcaagagcca	tttaaaaatc	tgaaaacagg	aaaatatgca	agaatgaggg	gtccccacac	3180
taatgtatgt	aaacaattaa	cagaggcagt	gcaaaaaata	accacagaaa	gcatagtaat	3240
atggggaaag	actcctaaat	ttaaactgcc	cataaaaaag	gaaacatgg	aaacatgg	3300
gacagagtat	tggcaagcca	cctggattcc	tgagtggag	tttgttaata	cccctccctt	3360

agtcaaattt	tggtaaccgt	tagagaaaga	accatagta	ggagcagaaa	ccttctatgt	3420
agatggggca	gctaacaggg	agactaaatt	aggaaaagca	ggatatgtt	ctaatacgagg	3480
aagacaaaaaa	gtgtcaccc	taactgacac	aacaaatcg	aagactgagt	tacaagcaat	3540
ttatctagct	ttgcaggatt	cgggattaga	agtaaacata	gtaacagact	cacaatatgc	3600
attaggaatc	attcaagcac	aaccagatca	aagtgaatca	gagttagtc	atcaaataat	3660
agagcagtta	ataaaaaaagg	aaaaggtcta	tctggcatgg	gtaccagcac	acaaaggaat	3720
tggagggaaat	gaacaagtag	ataaattagt	cagtgtgga	atcagggaaag	tactatttt	3780
agatggaata	gataaggccc	aagatgaaca	tgagaaat	cacagtaatt	ggagagcaat	3840
ggctagtgtat	tttaacctgc	cacctgttagt	agcaaaagaa	atagtagcca	gctgtgataa	3900
atgtcagcta	aaaggagaag	ccatgcatgg	acaagtagac	tgtgtccag	gaatatggca	3960
actagattgt	acacatttag	aaggaaaagt	tatcctggta	gcagttcatg	tagccagtgg	4020
atataatgaa	gcagaagttt	ttccagcaga	aacagggcag	gaaacagcat	attttctttt	4080
aaaatttagca	ggaagatggc	cagtaaaaac	aatacatact	gacaatggca	gcaatttcac	4140
cggtgctacg	gttagggccg	cctgttggtg	ggcggaatc	aagcaggaat	ttggatttcc	4200
ctacaatccc	caaagtcaag	gagtagtaga	atctatgaat	aaagaattaa	agaaaattat	4260
aggacaggtt	agagatcagg	ctgaacatct	taagacagca	gtacaaatgg	cagtattcat	4320
ccacaatttt	aaaagaaaaag	gggggattgg	gggttacagt	gcagggaaa	gaatagtaga	4380
cataatagca	acagacatac	aaactaaaga	attacaaaaa	caaattaca	aaattcaaaa	4440
tttcgggtt	tattacaggg	acagcagaaa	tccactttgg	aaaggaccag	caaagctcct	4500
ctggaaagg	gaagggcag	tagtaataca	agataatagt	gacataaaag	tagtgc当地	4560
aagaaaaagca	aagatcatta	gggattatgg	aaaacagatg	gcaggtgatg	attgtgtggc	4620
aagtagacag	gatgaggatt	agaacatgga	aaagtttagt	aaaacaccat	atgtatgttt	4680
cagggaaagc	tagggatgg	ttttatagac	atcaactatga	aagccctcat	ccaagaataa	4740
gttcagaagt	acacatccc	ctagggatg	ctagatttgt	aataacaaca	tattgggtc	4800
tgcatacagg	agaaagagac	tggcatttgg	gtcaggagt	ctccatagaa	tggagggaaa	4860
agagatata	cacacaagta	gaccctgaac	tagcagacca	actaattcat	ctgtattact	4920
ttgactgttt	ttcagactct	gctataagaa	aggccttatt	aggacacata	gttagcccta	4980
ggtgtgaata	tcaagcagga	cataacaagg	tagatctct	acaatacttg	gcactagcag	5040
cattaataac	acccaaaaaaag	ataaagccac	ctttgcctag	tgttacgaaa	ctgacagagg	5100
atagatggaa	caagccccag	aagaccaagg	gccacagagg	gagccacaca	atgaatggac	5160
actagagctt	ttagaggagc	ttaagaatga	agctgtttaga	cattttccct	ggatttggct	5220
ccatggctta	gggcaacata	tctatgaaac	ttatgggat	acttggcag	gagtggaaagc	5280

cataataaga attctgcaac aactgctgtt tatccatTTT cagaattggg tgTCGACATA 5340  
 gcagaatagg cgTTactcga cagaggagag caagaaatgg agccAGTAGA tcctAGACTA 5400  
 gagCCCTGGA AGCATCCAGG aAGTCAGCCT AAAACTGCTT gtACCAATTG CTATTGTA 5460  
 aAGTGTGCT tTCATTGCCA AGTTGTTTC ATAACAAAAG CCTTAGGCAT CTCCTATGGC 5520  
 aggaAGAAGC ggAGACAGCG acGAAGAGCT catcAGAACA gTCAGACTA tCAAGCTTCT 5580  
 CTATCAAAGC AGTAAGTAGT acATGTAATG caACCTATAc CAATAGTAGC AATAGTAGCA 5640  
 ttagtagtag caATAATAAT agCAATAGTT gtGTGGTCCA tagTAATCAT agAAATATAGG 5700  
 AAAATATTAA gACAAAGAAA aATAGACAGG tTAATTGATA gACTAAATAGA aAGAGCAGAA 5760  
 gACAGTGGCA ATGAGAGTGA aggAGAAATA tcAGCACTTG tGGAGATGGG gGTGGAGATG 5820  
 gggcaccatg ctccTTGGGA tGTTGATGAT ctGTAGTGCT acAGAAAAT tGTGGGTcAC 5880  
 agtctattat gggGTACCTG tGTGGAAAGGA AGCAACCACC ACTCTATTT gtGCATCAGA 5940  
 tgCTAAAGCA tATGATAACAG agGTACATAA tGTTTGGGCC acACATGCCT gtGTACCCAC 6000  
 agACCCCAAC ccACAAGAAG tagTATTGGT AAATGTGACA gAAAATTAA acATGTGGAA 6060  
 aaATGACATG gtAGAACAGA tGCAcTAGGA tATAATCAGT ttATGGGATC aaAGCCTAA 6120  
 gCCATGTGTA AAATTAACCC cACTCTGTGT tagTTAAAG tGCACTGATT tGAAGAATGA 6180  
 tactaatacc aATAGTAGTA gCGGGAGAAt gATAATGGAG aaAGGAGAGA tAAAAAAACTG 6240  
 ctcttcaat atcAGCACAA gcATAAGAGG tAAAGGTGcAG AAAGAAATATG cATTtttTA 6300  
 taaACTTGTAT ATAATACAA tagATAATGA tactACCAGC tATAAGTTGA caAGTTGTAA 6360  
 cacCTCAGTC attACACAGG CCTGTCCAAA gGTATCCTT gagCCAAATC ccATACATTa 6420  
 ttGTGCCCCG gCTGGTTTG cgATTCTAAA atGTAATAAT aAGACGTCA atGGAACAGG 6480  
 accATGTACA aATGTcAGCA cAGTACAATG tacACATGGA attAGGCCAG tagTATCAAC 6540  
 tcaACTGCTG tAAATGGCA gtCTAGCAGA agAAAGAGGTa gTAATTAGAT ctGTCAATT 6600  
 cacGGACAAT gCTAAAACCA taATGTACA gCTGAACACA tCTGTAGAAA tTAATTGTAC 6660  
 aAGACCCAAc aACAATAACAA gAAAAGAAt ccGTATCCAG agAGGACCAG ggAGAGCATT 6720  
 tGTTACAATA gGAAAATAG gAAATATGAG aCAAGCACAT tGTAACATTA gTAGAGCAAA 6780  
 atGGAATAAC actTTAAAAC agATAGCTAG caAAATTAAGA gAACAAATTG gAAATAATAA 6840  
 aACAATAATC ttTAAGCAAT CCTCAGGAGG ggACCCAGAA attGTAACGC acAGTTTAA 6900  
 ttGTGGAGGG gaATTtttCT actGTAAATTc AACACAACTG ttTAATAGTA ctTGGTTAA 6960  
 tagTACTTGG agTACTGAAG gGTCAAATAA cACTGAAGGA agTGACACAA tCACCCtCCC 7020  
 atGCAGAAATA aAACAAATTa tAAACATGTG gCAGAAAGTA ggAAAAGCAA tGTATGCC 7080  
 tCCCATCAGT ggACAAATTa gATGTTcATC AAATATTACA gGGCTGCTAT tAACAAAGAGA 7140  
 tGGTGGTAAT agCAACAAATG agTCCGAGAT CTTCAgACCT ggAGGAGGAG atATGAGGGA 7200

caattggaga agtgaattat ataaatataa agtagtaaaa attgaaccat taggagtagc	7260
acccaccaag gcaaagagaa gagtggtgca gagagaaaaa agagcagtgg gaataggagc	7320
tttgcctt gggttcttgg gagcagcagg aagoactatg ggccgcgcct caatgacgct	7380
gacggtacag gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgctgag	7440
ggctattgag gcgcaacagc atctgttgc actcacagtc tggggcatca agcagctcca	7500
ggcaagaatc ctggctgtgg aaagataacct aaaggatcaa cagtcctgg ggatttgggg	7560
ttgctctgga aaactcattt gcaccactgc tgtgccttgg aatgcttagt ggagtaataa	7620
atctctggaa cagatttggaa atcacacgac ctggatggag tgggacagag aaattaacaa	7680
ttacacaagc ttaatacact ccttaattga agaatcgcaa aaccagcaag aaaagaatga	7740
acaagaatta ttggaatttag ataaatggc aagtttgtgg aattggtttta acataacaaa	7800
ttggctgtgg tatataaaat tattcataat gatagtagga ggcttggtag gttaagaat	7860
agttttgct gtacttcta tagtgaatac agttaggcag ggatattcac cattatcggt	7920
tcagaccac ctcccaaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg	7980
tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg	8040
ggacgatctg cggagcctgt gcctttcag ctaccaccgc ttgagagact tactcttgat	8100
tgttaacgagg attgtggAAC ttctggac cagggggtgg gaagccctca aatattggtg	8160
gaatctccta cagtatttggaa gtcaggaact aaagaatagt gctgttagt tgctcaatgc	8220
cacagccata gcagtagctg agggacaga tagggttata gaagtagtac aaggagcttgc	8280
tagagctatt cgccacatac ctagaagaat aagacagggc ttggaaagga ttttgctata	8340
agatgggtgg caagtggtaaaaatg tgattggatg gcctactgta agggaaagaa	8400
tgagacgagc tgagccagca gcagataggg tggagcagc atctcgagac ctggaaaaac	8460
atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttgccttcc tggctagaag	8520
cacaagagga ggaggagggtg ggtttccag tcacacctca ggtaccttta agaccaatga	8580
cttacaaggc agctgttagat cttagccact tttaaaaga aaaggggggc ctggaaaggc	8640
taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct	8700
acttccctga ttacgagaac tacacaccag ggccagggtt cagatatcca ctgaccttgc	8760
gatggtgcta caagcttagta ccagttgagc cagataagat agaagaggcc aataaaggag	8820
agaacaccag cttgttacac cctgtgagcc tgcattggat ggatgaccgc gagagagaag	8880
tgttagatgttggac agccgcctag catttcataca cgtggccgc gagctgcatac	8940
cggagtactt caagaactgc tgacatcgag cttgtacaa gggactttcc gctggggact	9000
ttccaggggag gcgtggcctg ggcgggactg gggagtggcg agccctcaga tcctgcataat	9060
aagcagctgc tttttgcctg tactgggtct ctctggtag accagatctg agcctggag	9120

ctctctggct aactaggaa cccactgctt aagcctcaat aaagcttgc	ttgagtgc	9180
c		9181
<210> 15		
<211> 29		
<212> RNA		
<213> Homo sapiens		
<400> 15		
ggcagaucug agccugggag cucucugcc		29
<210> 16		
<211> 52		
<212> RNA		
<213> Homo sapiens		
<400> 16		
uuuuuuuaggg aagaucuggc cuuuccuacaa gggaggcca gggauuuuuc uu		52
<210> 17		
<211> 9413		
<212> DNA		
<213> Homo sapiens		
<400> 17		
ttggggcga cactccacca tagatcactc ccctgtgagg aactactgtc ttcacgcaga		60
aagcgctctag ccatggcggtt agtatgagtg ttgtgcagcc tccaggaccc cccctcccg		120
gagagccata gtggctcgca gaaaccggta gtacaccgga attgccagga cgaccgggtc		180
ctttcttggta tcaacccgct caatgcctgg agatttggc gtgcccccgca gagactgcta		240
gccgagtagt gttgggtcgc gaaaggcctt gtggtaactgc ctgataggtgttgcgagt		300
gccccgggag gtctcgtaga ccgtgcata tgacacaaa tcctaaacct caaaagaaaaaa		360
ccaaacgtaa caccaaccgc cgccccacagg acgttaagtt cccggggcgtt ggtcagatcg		420
ttgggtggagt ttacctgttgcg cccgcgcagg gcccagggtt ggggtgtgcgc ggcacttagga		480
agacttccga gcggtcgca cctcgtggaa ggcgacaacc tatccccaaag gctcgccggc		540
ccgagggttag gacctggct cagccccgggtt acccttggcc cctctatggc aacgagggtta		600

tgggggtggc	aggatggctc	ctgtcacccc	gtggctctcg	gcctagttgg	ggccccacag	660	
accccccggcg	taggtcgctgt	aatttggta	aggtcatcga	tacccttaca	tgcggcttcg	720	
ccgacacctat	ggggtacatt	ccgcttgcg	gccccccct	agggggcgt	gccagggccc	780	
tggcacatgg	tgtccgggtt	ctggaggacg	gcgtgaacta	tgcaacaggg	aatctgccc	840	
gttgctcttt	ctctatcttc	ctcttagtt	tgctgtcttg	tttgaccatc	ccagcttccg	900	
cttacgaggt	gcgcaacgtg	tccggatat	accatgtcac	gaacgactgc	tccaactcaa	960	
gtattgtgt	tgaggcagcg	gacatgatca	tgcacacccc	cgggtgcgtg	ccctgcgtcc	1020	
gggagagtaa	tttctcccg	tgctggtag	cgctcactcc	cacgctcg	gccaggaaca	1080	
gcagcatccc	caccacgaca	atacgacgcc	acgtcgattt	gctcg	ggctgctgctc	1140	
tctgtccgc	tatgtacgtt	ggggatctct	gccccgt	ttttctcg	tcccg	1200	
tcaccttctc	acctcgccgg	tatgagacgg	tacaagattt	caattgtca	atctatcccg	1260	
gccacgtatc	aggtcaccgc	atggcttggg	atatgtat	gaactggtca	cctacaacgg	1320	
ccctagtggt	atcgcagcta	ctccggatcc	cacaaggcg	cgtggacatg	gtggcggggg	1380	
cccaactgggg	tgtcttagcg	ggccttg	actattccat	ggtggggaa	tggctaagg	1440	
tctttagttt	gatgtactc	tttgc	ttgacggca	cacccacgt	acagggggaa	1500	
ggtaggcctc	cagcacccag	agcctcg	cctgg	tctcaca	aaaa	1560	
tccaactcg	gaacaccaac	ggcagtc	acatcaacag	gaccg	ctcg aatg	1620	
actccctcca	aactgggttc	attgc	tgttctacgc	acacagg	ttc aacgcgtcc	1680	
ggtgc	ccaga	gcatggct	agctg	ccatcgatg	gttcg	ctcg aggtgg	1740
ccatca	tgc	atgc	tttgc	acatcgatg	gttcg	ctcg ttttgc	1800
ctcgaccgt	cg	ggatcg	cctgc	agg	gttgc	ctcg ttttgc	1860
cgagccctgt	tgt	tagtgggg	acgaccg	at	ttcggc	tcctacgt	1920
agaatgagac	agacgt	gtct	ctacttag	ac	acg	ccgtgcaaca	1980
ggtgcacgt	gat	gac	actgg	cca	acgt	cgggggcc	2040
tcgggggggt	cg	gaaca	ac	cttgg	cc	ccgtgcaaca	2100
aggccactta	caca	aa	gtgt	ggc	cc	ccgtgcaaca	2160
accatacag	gct	ctgg	ca	ccct	cc	ccgtgcaaca	2220
tgtatgtgg	gg	ggcgtgg	ca	agg	cc	ccgtgcaaca	2280
gctgtactt	gg	aggacagg	gat	agg	cc	ccgtgcaaca	2340
agtggcagat	act	gccc	ct	ttc	cc	ccgtgcaaca	2400
atcttcacccg	ga	acatcg	tg	ca	cc	ccgtgcaaca	2460
ccttgcaat	caa	atggg	tg	atgg	cc	ccgtgcaaca	2520

tctgtgcctg	cttgtggatg	atgctgctga	tagcccaggc	tgaggccacc	ttagagaacc	2580
tggtgtcct	caatgcggcg	tctgtggccg	gagcgcatgg	ctttctctcc	ttcctcgtgt	2640
tcttctgcgc	cgcctggta	atcaaaggca	ggctggtccc	tggggcggca	tatgctctct	2700
atggcgtatg	gccgttgctc	ctgctcttc	tggccttacc	accacgagct	tatgccatgg	2760
accgagagat	ggctgcatcg	tgcggaggcg	cggttttgt	aggctggta	ctcttgacct	2820
tgtcaccata	ctataaggtg	ttcctcgcta	ggctcatatg	gtggttacaa	tattttatca	2880
ccagagccga	ggcgcacttg	caagtgtggg	tccccctct	aatgttcgg	ggaggccgcg	2940
atgccccatcat	cctccttaca	tgcgcggtcc	atccagagct	aatcttgcac	atcaccaaac	3000
tcctgctcgc	catactcggt	ccgctcatgg	tgctccaggc	tggcataact	agagtgccgt	3060
actttgtacg	cgctcagggg	ctcatccgt	catgcatttt	agtgcggaaag	gtcgctggag	3120
gccactatgt	ccaaatggcc	ttcatgaagc	tggccgcgt	gacaggtacg	tacgtatatg	3180
accatcttac	tccactgcgg	gattggggcc	acgcgggcct	acgagacatt	gcggtggcag	3240
tagagccgt	cgtcttctct	gacatggaga	ctaaactcat	cacctggggg	gcagacaccg	3300
cggcgtgtgg	ggacatcatc	tcgggtctac	cagtcctcgc	ccgaaggggg	aaggagatac	3360
ttcttaggacc	ggccgatagt	tttggagagc	aggggtggcg	gctccttgcg	cctatcacgg	3420
cctatttcca	acaaacgcgg	ggcctgcttgc	gctgtatcat	cactagcctc	acaggtcggg	3480
acaagaacca	ggtcgatggg	gaggttcagg	tgctctccac	cgcaacgc当地	tctttcttgg	3540
cgacctgcgt	caatggcgtg	tgttggaccg	tctaccatgg	tgccggctcg	aagaccctgg	3600
ccggcccgaa	gggtccaatc	acccaaatgt	acaccaatgt	agaccaggac	ctcgctggct	3660
ggccggcgcc	ccccggggcg	cgctccatga	caccgtgcac	ctgcggcagc	tcggaccttt	3720
acttggtcac	gaggcatgct	gatgtcggttc	cggtgcgcgc	gcggggcgac	agcaggggg	3780
gcctgctttc	ccccaggccc	atctcctacc	tgaagggctc	ctcggttgg	ccactgcttt	3840
gcccttcggg	gcacgttgta	ggcatcttcc	ggctgctgt	gtgcacccgg	gggttgc当地	3900
aggcgggtgg	cttcatacc	gttgagttct	tggaaactac	catgcggct	ccggtcttca	3960
cagacaactc	atcccctccg	gccgtaccgc	aaacattcca	agtggcacat	ttacacgctc	4020
ccactggcag	cgcaagagc	accaaagtgc	cggtgcata	tgcagccaa	gggtacaagg	4080
tgctcgctt	aaacccgtcc	gttgcccca	cattggctt	tggagctat	atgtccaagg	4140
cacatggcat	cgagccta	atcagaactg	ggtaaggac	catcaccacg	ggcggcccca	4200
tcacgtactc	cacctattgc	aagtcccttgc	ccgacgggtgg	atgcctccgg	ggcgctatg	4260
acatcataat	atgtgatgaa	tgccactcaa	ctgactcgac	taccatcttgc	ggcatcgca	4320
cagtcccttgg	tcaggcagag	acggctggag	cgcggctcg	cgtgctcgcc	accgcccacgc	4380
ctccgggatc	gatcaccgtg	ccacacccca	acatcgagga	agtggccctg	tccaacactg	4440

gagagattcc	cttcttatggc	aaagccatcc	ccattgaggc	catcaagggg	ggaaggcattc	4500
tcatcttctg	ccattccaag	aagaagtgtg	acgagctcgc	cgcaaagctg	acaggcctcg	4560
gactcaatgc	tgtagcgtat	taccggggtc	tcgatgtgtc	cgtcataccg	actagcggag	4620
acgtcgttgt	cgtggcaaca	gacgctctaa	tgacgggtt	taccggcgcac	tttgactcag	4680
tgatcgactg	caacacatgt	gtcacccaga	cagtcgattt	cagttggat	cccacccatca	4740
ccattgagac	gacaacgctg	ccccaaagacg	cggtgtcgcg	tgcgcagcgg	cgaggttagga	4800
ctggcagggg	caggagtggc	atctacaggt	tttgtgactcc	aggagaacgg	ccctcaggca	4860
tgttcgactc	ctcggtcctg	tgtgagtgtct	atgacgcagg	ctgcgccttgg	tatgagctca	4920
cgcggctga	gacctcggtt	aggttgcggg	cttacctaaa	tacaccaggg	ttgcccgtct	4980
gccaggacca	cctagagttc	tgggagagcg	tcttcacagg	cctcaccac	atagatgccc	5040
acttcttgc	ccagacccaaa	caggcaggag	acaacccc	ctacctggta	gcataccaag	5100
ccacagtgtg	cgcagggtct	caggctccac	ctccatcg	ggaccacatg	ttgaaagtgtc	5160
tcatacggt	aaagccccaca	ctgcatggc	caacgccc	ctgtacagg	ctaggagccg	5220
ttcaaaatga	ggtcaactctc	acacacccca	taaccaaata	catcatggca	tgcatgtcg	5280
ctgaccttgg	ggtcgtca	agcacctggg	tgcttaggt	cgagtcctt	gcccgtctgg	5340
ccgcgtactg	cctgacgaca	ggcagcgtgg	tcattgtggg	caggatc	ttgtccggg	5400
ggccagctgt	tattcccgac	agggaaagtcc	tctaccagga	tttcgatgag	atgaaagagt	5460
gtgcttcaca	cctcccttac	atcgagcaag	aatgcagct	cgccgagcaa	ttcaaacaga	5520
aggcgctcgg	attgctgcaa	acagccacca	agcaagcgg	ggctgctgt	cccgtgg	5580
agtccaaatg	gcgagccctt	gaggcttct	gggcgaaaca	catgtggaa	ttcatcagcg	5640
ggatacagta	cttggcagggc	ctatccactc	tgccctggaaa	ccccgcgata	gcatcattga	5700
tggctttac	agccctatc	accagccgc	tcaccacccca	aaataccctc	ctgttaaca	5760
tcttgggggg	atgggtggct	gcccaactcg	ctccccccag	cgctgcttc	gtttcgtgg	5820
gcccggcat	tgccgggtcg	gccgttggca	gcataagg	cggttggata	tttgtggaca	5880
ttctggcggg	ctatgggcg	gggggtggct	gcmcactcg	ggcctttaag	gtcatgagcg	5940
gcgcgcgc	ctccactgag	gatctggta	atttactccc	tgccatc	tctcctggcg	6000
ccctgggtgt	cggggtcg	tgccgcagca	tactgcgtcg	gcacgtggc	ccgggagagg	6060
gggctgtgca	gtggatgaac	cggctgatag	cggtcgcttc	gcggggtaac	cacgtctccc	6120
ccacgcacta	tgtgcccggag	agcgacgccc	cggcgcgtgt	tactcagatc	ctctccagcc	6180
ttaccatcac	tcaatggctg	aaggaggctt	atcagtgat	taatgaggac	tgctccacgc	6240
cttgttccgg	ctcggtggcta	aaggatgttt	gggactggat	atgcacggtg	ttgagtgact	6300
tcaagacttg	gctccagtc	aagtcctgc	cgcggtaacc	ggacatccct	ttccgtcat	6360

gccaacgcgg	gtacaaggga	gtctggcggg	gggatggcat	catgcaaacc	acctgcccatt	6420
gtggagcaca	gatcaccggaa	catgtcaaaa	atggctccat	gaggattgtt	gggc当地	6480
cctgcagcaa	cacgtggcat	ggaacattcc	ccatcaacgc	atacaccacg	ggccc当地	6540
cgc当地	agcgc当地	tattccaggg	cgctgtggcg	ggtggctgct	gaggagtacg	6600
tggaggttac	gc当地	gatccact	acgtgacggg	catgaccact	gacaacgtga	6660
aatgcccatt	ccaggttcca	gccc当地	ttttcacgga	ggtggatgga	gtacggttgc	6720
acaggtatgc	tccagtggtgc	aaacctctcc	tacgagagga	ggtcgtattc	caggtc当地	6780
tcaaccagta	cctggtc当地	tcacagctcc	catgtgagcc	c当地	gtggc当地	6840
tcacttccat	gctcaccgc当地	ccctctcata	ttacagcaga	gacggccaag	c当地	6900
ccagggggtc	tccccc当地	ttggccagct	ctttagctag	ccagttgtct	gc当地	6960
tgaaggcgac	atgtactacc	catcatgact	ccccggacgc	tgacctcatc	gaggccaacc	7020
tcctgtggcg	gc当地	gaggagatg	ggc当地	tcaccgtgt	ggagtc当地	7080
taatccctgga	ctctttcgat	ccgattc当地	cggtggagga	tgagagggaa	atatccgtcc	7140
cgccggagat	c当地	ccaggaagt	tccccc当地	gttgc当地	tggc当地	7200
cggattacaa	ccctccactg	ctagactcct	ggaaggaccc	ggactacgtc	ccccgggtgg	7260
tacacgggtg	ccctttgcca	tctaccaagg	cccccccaat	accacctcca	cgaggaaga	7320
ggacggttgc	c当地	tccaccgtgt	cttctgc当地	ggc当地	gctactaaga	7380
cctttggcag	ctccgggtcg	tc当地	acagcggcac	ggc当地	c当地	7440
aggcctccga	cgacggcgac	aaaggatccg	acgttgagtc	gtactc当地	atgccccccc	7500
tcgagggaga	gc当地	cccgacctca	gca当地	ttggtctacc	gtgagc当地	7560
aagctggta	ggacgtcg	tgctgctcaa	tgtcc	tatac	atggacagg	7620
c当地	tgccgaggag	agcaagttgc	ccatcaatcc	gtt当地	gcaac	7680
gtcaccacag	tatggtctac	tccacaacat	ctcgacgc	aagtctgc当地	cagaagaagg	7740
tcaccttga	cagactgcaa	gtcc	tgac	accactaccg	ggacgtgctc	7800
aggcgaaggc	gtcc	ca	aggcttaggc	ttctatctat	agaggaggcc	7860
c当地	ttcgccaaa	tccaaatttgc	gtacggggc	gaaggacgtc	cgaggc当地	7920
ccagcaggc当地	cg	caaccac	atccgctccg	tgtggagga	ctt当地	7980
caccaattga	taccaccatc	atggcaaaa	atgagg	tttctgctccaa	ccagagaaag	8040
gaggccgcaa	gc当地	ctt当地	tcc	ccagac	gtatgc当地	8100
agatggccct	ttacgacgt	gtctcc	acc	gggtacgt	gtatgc当地	8160
gattccagta	ctctcctggg	cagcgggtcg	agttcc	tgg	aaatcaaaga	8220
aatgc当地	ggc当地	tctca	tatgacaccc	gctgcttga	ctcaacggc	8280

acatccgtac	tgaggaatca	atttaccaat	gttgtgactt	ggcccccggaa	gccaggcagg	8340
ccataaggc	gctcacagag	cggtttatg	tcgggggtcc	cctgactaat	tcgaaggggc	8400
agaactgcgg	ttatcgccgg	tgccgccaa	gtggcgtgct	gacgactagc	tgccgcaaca	8460
ccctcacatg	ttacttgaag	gccactgcgg	cctgtcgagc	tgcaaagctc	caggactgca	8520
cgtgctcgt	gaacggagac	gaccctgtcg	ttatctgtga	gagtgcggga	acccaggagg	8580
atgcggcggc	cctacgagcc	ttcacggagg	ctatgactag	gtattccgccc	ccccccgggg	8640
acccgccccca	accagaatac	gacttggagc	tgataacgtc	atgctcctcc	aatgtgtcgg	8700
tcgcgcacga	tgcacccggc	aaaagggtgt	actacctcac	ccgtgacc	accacccccc	8760
tcgcacgggc	tgcgtggag	acagtttagac	acactccagt	caactctgg	ctaggcaata	8820
tcatcatgt	tgcgcacc	ctatggcga	ggtatgattct	gatgactcat	ttcttctcta	8880
tccttctagc	tcaggagcaa	cttggaaaag	ccctggattt	tcaatctac	ggggcctgtt	8940
actccattga	gccacttgac	ctacccatcaga	tcattgaacg	actccatgg	cttagcgcac	9000
tttcactcca	cagttactct	ccaggtgaga	tcaatagggt	ggcttcatgc	ctcagaaac	9060
ttgggttacc	gccttgcga	gtctggagac	atcggccag	aagtgtccgc	gctaagctac	9120
tgtcccaggg	ggggagggct	gccacttgcg	gcaagtacct	cttcaactgg	gcagtaaaga	9180
ccaagcttaa	actcactcca	atccggctg	cgtcccagct	agacttgc	ggctggttcg	9240
ttgctggta	caacggggga	gacatatac	acagcctgtc	tcgtgc	ccccgttgg	9300
tcatgttgc	cctactccta	ctttctgtag	gggttaggcat	ctacctgctc	cccaaccgg	9360
gaacggggag	ctaaccaactc	caggccaaata	ggccattccc	ttttttttt	ttc	9413

&lt;210&gt; 18

&lt;211&gt; 328

&lt;212&gt; RNA

&lt;213&gt; Homo sapiens

<400> 18	uuggggcga	cacuccacca	uagaucacuc	ccugugagg	aacuacuguc	uucacgcaga	60
	aaggcucuag	ccauuggcguu	aguauugagug	uugugcagcc	uccaggaccc	ccccucccg	120
	gagagccaua	guggucugcg	gaaccgguga	guacaccgga	auugccagga	cgaccggguc	180
	cuuucuugga	ucaacccgcu	caauggcugg	agauuuggc	gugcccccgc	gagacugcua	240
	gccgaguagu	guugggucgc	gaaaggccuu	gugguacugc	cugauagggu	gcuuggcagu	300
	ccccgggag	gucucguaga	ccgugcau				328

<210> 19

<211> 14

<212> RNA

<213> Homo sapiens

<400> 19

auuugggcgu gccc

14

<210> 20

<211> 27

<212> RNA

<213> Homo sapiens

<400> 20

gccgaguagu guugggucgc gaaaggc

27

<210> 21

<211> 340

<212> DNA

<213> Homo sapiens

<400> 21

atgggcggag ggaagctcat cagtgggcc acgagctgag tgcgtcctgt cactccactc 60

ccatgtccct tgggaaggc tgagactagg gccagaggcg gccctaacag ggctctccct 120

gagcttcagg gaggtgagtt cccagagaac ggggctccgc gcgaggtcag actgggcagg 180

agatgccgtg gaccccgccc ttcggggagg ggcccgccgg atgcctcctt tgccggagct 240

tggAACAGAC tcacggccag cgaagtgagt tcaatggctg aggtgaggtt ccccgccagg 300

gacctataa cccaaattcag accactctcc tccggccatt 340

<210> 22

<211> 349

<212> DNA

<213> Homo sapiens

<400> 22

gagggaaagtc cgggctcaca cagtctgaga tgattgtagt gttcgtgctt gatgaaacaa 60  
 taaatcaagg cattaatttg acggcaatga aatatcctaa gtctttcgat atggatagag 120  
 taatttggaa gtgccacagt gacgttagct ttatagaaat ataaaagggtg gaacgcggta 180  
 aaccctcga gtgagcaatc caaatttggt aggagcactt gtttaacgga attcaacgta 240  
 taaacgagac acacttcgca aatgaagtg gtgtagacag atggttatca cctgagttacc 300  
 agtgtgacta gtgcacgtga ttagtacgt ggaacagaac gcggcttat 349

<210> 23

<211> 377

<212> DNA

<213> Homo sapiens

<400> 23  
 gaagctgacc agacagtcgc cgcttcgtcg tcgtcctctt cgggggagac gggcggaggg 60  
 gagggaaagtc cgggctccat agggcagggt gccaggtaac gcctgggggg gaaaccac 120  
 accagtgcaa cagagagcaa accgcccgtg gcccgccaa gcgggatcag gtaagggtga 180  
 aagggtgcgg taagagcgca ccgcgcggct gtaacagtc cgtggcacgg taaactccac 240  
 ccggagcaag gccaaatagg gttcataag gtacggcccg tactgaaccc ggtaggctg 300  
 cttgagccag tgagcgttgc ctggcctaga tgaatgactg tccacgacag aaccggctt 360  
 atcggtcagt ttcacct 377

<210> 24

<211> 38110

<212> DNA

<213> Homo sapiens

<400> 24  
 ccaccggta ccatcttgcc gaccatggcc ccacaatagg gcccggaga cccggcgtca 60  
 gtggtggcg gcacggtcag taacgtctgc gcaacacggg gttgactgac gggcaatatc 120  
 ggctccatag cgtcgccgc ggatacagta aaggagcatt ctgtgacgga aaagacgccc 180  
 gacgacgtct tcaaacttgc caaggacgag aaggtcgaat atgtcgacgt ccgggtctgt 240  
 gacctgcctg gcatcatgca gcacttcacg attccggctt cggccttga caagagcgtg 300  
 tttgacgacg gcttggcctt tgacggctcg tcgattcgca ggttccagtc gatccacgaa 360  
 tccgacatgt tgcttcttcc cgatcccgag acggcgcgca tcgacccgtt ccgcgcggcc 420

aagacgctga atatcaactt	480
ctttgtcac gaccgttca	
ccctggagcc gtactccgc	
gaccgcgcac acatcgcccc	540
caaggccgag aactacctga	
tcagcactgg catcgccgac	
accgcatact tcggcgccga	600
ggccgaggttc tacatttcg	
attcggtgag ctgcactcg	
cgcgcacacg gctccttcta	660
cgaggtggac gccatctcg	
ggtggtggaa caccggcg	
gcgaccgagg cgcacggcag	720
tcccaaccgg ggctacaagg	
tccgccacaa gggcgggtat	
ttcccagtgg cccccaacga	780
ccaatacgtc gacctgcgcg	
acaagatgct gaccaacctg	
atcaactccg gttcatacct	840
ggagaaggc caccacgagg	
tggcagcgg cgacaggcc	
gagatcaact accagttcaa	900
ttcgctgctg cacgcccc	
acgacatgca gttgtacaag	
tacatcatca agaacaccgc	960
ctggcagaac ggcaaaacgg	
tcacgttcat gcccaagccg	
ctgttcggcg acaacgggtc	1020
cggcatgcac tgtcatcgt	
cgctgtggaa ggacggggcc	
ccgctgatgt acgacgagac	1080
gggttatgcc ggtctgtcg	
acacggcccg tcattacatc	
ggcggcctgt tacaccacgc	1140
gccgtcgctg ctggccttca	
ccaacccgac ggtgaactcc	
tacaagcggc tggttcccg	1200
ttacgaggcc ccatcaacc	
tggtctatacg ccagcgcaac	
cggtcggcat ggtgcgcac	1260
cccgatcacc ggcagcaacc	
cgaaggccaa gggctggag	
ttccgaagcc ccgactcgac	1320
ggcaaccgg tatctggcgt	
tctcgccat gctgatggca	
ggcctggacg gtatcaagaa	1380
caagatcgag cgcaggcgc	
ccgtcgacaa ggatctctac	
gagctgccgc cggaaagaggc	1440
cgcgagtatc cgcagactc	
cgaccagct gtcaagatgt	
atcgaccgtc tcgaggccga	1500
ccacgaatac ctcaccgaag	
gaggggtgtt cacaacacac	
ctgatcgaga cgtggatcag	1560
tttcaagcgc gaaaacgaga	
tcgagccggt caacatccgg	
ccgcattccct acgaattcgc	1620
gctgtactac gacgttaag	
gactcttcgc agtccgggt	
tagagggagc ggcgtgtcgt	1680
tgccaggccg ggcgtcgagg	
tttttcgtt ggtgacgggt	
gccggcaacg gcgccgcac	1740
caccgctcg aagagccgt	
ttaagaacgt tcaaggacgt	
ttcagccggg tgccacaacc	1800
cgcttggcaa tcatctcccg	
accgcccggc gggttgtt	
tcacatcgac cgaactcaa	1860
gccacgtcgt cgcccaggcg	
tgtcgctcg gcccgtttag	
gttaagtgtc gggattcgt	1920
cgtgcggccg ggcgtccacg	
ctgaccaacg gggcagtcaa	
ctccccgaaca ctttgcgcac	1980
taccgcctt gcccggccg	
tcacccgtt gtagttgtcc	
aggaattccc caccgtcgac	2040
gtttcgccag cggccggcga	
ccgcgaccgc attgagctgg	
cgcggccgtc cccgcggct	2100
gtcgggtggc ttggccgcga	
ccaacaccag cgcgttgcgg	
gcccgggtgg cggtcagcca	2160
ggcctgacgg agcagctcca	
cgtcggctgc gggaaaccaga	
tcggccggccg cgatgacatc	2220
caggattgc agcgtcgagg	
tgttgtgcag ggcgggaacc	
tggtgcgcac gctgtactgc	2280
cagcaactgc acggccatt	
cgatgtcgac cagtccggc	
cggcccaagtt tggtgtgtgt	2340
gttggggtcg gcaccgcgcg	
gcaaccgctc ggactcgata	

cgggccttga	tgccggcgaat	ctcgccgacc	gagtcagcgg	acacaccgtc	ggcgat	2400
cgcgtttgt	cgaccatccg	taggaatcgc	tgacccaact	cggcatcgcc	ggcaaccgcg	2460
tgtgcgcgt	gcagggcctg	gatctccat	ggctgtgccc	actgctcgta	gtatgcggcg	2520
taggacccc	gggtgcggac	cagcggaccg	ttgcggccct	cgggtcgcaa	attggcgtcg	2580
agctccagcg	gcccgtcgac	gctgggtgtc	cccagcagcg	cccgaacccg	ctcgccgatc	2640
gatgtcgacc	atttcaccgc	ccgtgcac	tcgacgcccgg	tggccggctc	acagacgaac	2700
atcacgtcg	catccgaccc	gtagccaaac	tcggcaccac	ccagccgacc	catgcccgt	2760
accgcgatgg	ccgcgggggc	gcgatcg	tcgggaaggc	tggccggat	catgacgtcc	2820
agcgcggcct	gcagcaccgc	cacccacacc	gacgtcaacg	ccggcacac	ctcggtgacc	2880
tcgagcaggc	cgagcagg	cgccgaaccg	atgcggccca	gctctcgacg	acgcagcgt	2940
cgcgcggcgg	cgatggcccg	ctccgggtcg	ggtagcggc	tcgcccggc	gatcagcgc	3000
cgagccacgg	cgccggggtc	ggtctcg	agcttcggc	ccgcaggccc	gtcctcg	3060
tgctggatga	cccgccggcgc	gcgcataac	agatccggca	catacggca	gtacccaag	3120
acatgcatga	gccgcttggc	caccgcgggc	ttgtcccgca	gcgtggccag	gtaccagtt	3180
tcgggtggcca	gcgcctca	gagccgcgg	taggcagca	gtccgcgtc	ggatcgggg	3240
gcatacgaca	tccagtc	cagcctggc	agcagcaccg	actgcacccg	tccgcggcgg	3300
ccgctttgat	tgaccaacgc	cgacatgtgt	ttcaacgcgg	tctgcggtcc	ctcgtagccc	3360
agcgcggcca	gccggcgccc	cgcggcctcc	aacgtcatgc	cgtggcgat	ctccaacccg	3420
gtcgggcca	tcgattccag	cagcggt	tagaagagtt	tggtgtgtaa	cttcgacacc	3480
cgcacgttct	gttcttgag	ttcctccgc	agcacccgg	ccgcatacg	tccgc	3540
ggccggatgt	ggccgcgcg	cgcacccag	cgcactgc	cctcg	tccgc	3600
agcaggtgg	tgcgttgag	ccgctgcaac	tgcagtcggt	gctcg	gagc	3660
tcatacgacg	cgtcatgtt	cgcgcgtcc	tcaccccga	tgtagcc	ttcgccaa	3720
gccgccaatg	cgtccaccgt	ggacgcccacc	cgtaa	cgtcg	ctacg	3780
agctgcagta	gctgtacggc	gaactccacg	tcgcgcaatc	cgcgcgt	gagtt	3840
tcgcggccgc	ggacatcg	ggcacc	tgctccaccc	gccgcgc	catggc	3900
tcgaccacaa	agtcttcg	ctcg	cagg	ccacacca	tcggc	3960
taacgctcgc	caagttccgc	gtcgccaa	actggccgt	ctt	cagcaa	4020
tcccagg	tggcc	ctgg	tagtag	g	cgatgtcg	4080
tcccccgtgc	gccc	cc	acgc	agg	cgatccac	4140
acccgcac	tctcg	ctcg	cacgc	ggc	ggcg	4200
atgacatcg	cgtcg	ctcg	actcg	cgac	cgacttgc	4260

accgccaggc	gcgggtggcgg	gtgctcgccg	cacacgctcg	cctcgccac	gcgca	4320	-	
gccc	cagg	cggtccgc	ggcgtccgc	aggcgtgcgg	ccaccacgg	aatggcagc	4380	
accgg	ttcg	cgtcgt	cctcgaccgt	cgcgccagg	tcgagagcgg	ccagcattag	4440	
cggtact	ttcg	caatcg	gtgcacgagc	gagccggca	taccctccg	ttcc	4500	
cactcgac	ga	acgaccgctg	cagctggtca	tgggacggca	gtgtgac	ttt	4560	
aatttccagg	actgcggat	g	gcgaccagg	tgatcgcc	acgccc	cgagccc	4620	
accgagaaca	gccgccc	cg	cagactgcgt	tcgcgcagca	gagccgcgtt	gagctcg	4680	
catccgg	gt	ctggatt	tc	cgacagccgg	atcaaggcgc	gcagcgcggc	atcg	4740
ggagcgcgt	ac	agcgcacca	cagcagg	acgtgcgc	ct	gatcccac	4800	
cccagctg	g	ccagacg	ctc	accagcagg	gggtcaacta	atccgagccg	gccaac	4860
ggcaact	tcg	g	ccgctgcgt	ggcagttt	gtc	acgacca	cgacggtag	4920
tcggcgt	cg	atcaacc	gt	agatctggc	ta	acagcaca	ggtaggtgc	4980
ggcgt	gac	gt	ggctgcgg	ta	tcc	gtcg	caa	5040
aaaacgt	gt	ct	ccccaaagg	ctccgc	gac	agttcgagg	cctccat	5100
ctatccaa	ac	tg	gacgg	gg	ttc	tcggtac	ccgcgtc	5160
tcccata	ac	tg	cctc	ctg	cc	gtcg	gggtgtgagg	5220
cccgcgg	ca	gc	gacacgg	ga	at	gtcaga	taggatt	5280
acttcgac	cc	gc	ccgcgac	gg	t	tgttgc	acgccc	5340
gatcggtt	gg	cg	ggccccc	ca	g	tcgac	cccg	5400
ttgtaa	ag	tg	gacccact	tt	gt	tgacc	g	5460
ccggc	gat	ga	acgat	tt	tc	ccact	g	5520
gcgtt	gac	at	cac	cc	cc	tgcata	tcg	5580
ccgaat	gg	gt	ggc	at	cc	tcgaa	ccgg	5640
acgt	ac	tg	ggc	tc	cc	tcgac	gtcg	5700
atctc	ct	gt	ggc	tt	cg	tgt	gat	5760
tccagg	gg	ca	gtc	tc	gt	gtc	cg	5820
aaatag	cc	gt	gtc	gg	tt	gg	gt	5880
aggaaga	act	cg	at	tc	gg	tc	gg	5940
agctgc	cc	ca	tc	ac	cc	tc	cc	6000
atgtcg	aaa	ac	at	cc	cc	tc	gg	6060
tggaagg	tcg	ac	ggg	tc	gg	cc	ac	6120
tcgatcg	agg	at	cc	gtc	gg	cc	ac	6180

gcgatggcga ccgacttgag gaaaccgagc acgtctgtga accacagccg gacgaagcgg	6240
atgtcgcgtt cttccagggt acgaagaacg aattccttct gtcggtccat acctcgaaca	6300
gtatgcactg tctgttaaaa ccgtgttacc gatgcccggc cagaagcgtt gcggggcggc	6360
ccgcaagggg agtgcgcggt gagttcaggg cgccgaccgc agactcgtcg gcggcaaggt	6420
cccgtcgaga aaatagtgca tcaccgcaga gtccacacac tggttccat cgaacaccgc	6480
agtgtttgg gtgccgtcga aggtgatcag cggtgccccc agctggcggg ccaggtctac	6540
cccgactga tacggagtgg ccgggtcgtg ggtggtggac accacgacga cttgccagc	6600
cccgccggc gccgcggggt gcggcgtcga cttgccggc accggccaca gcgcgcacag	6660
atcgccccgg gcgatccgg tgaactgccc gtagctaagg aacggggcga cctgacggat	6720
ccgttgtcg gcggccaccc aggccgttgg atcggccgt gtgggcgcgt cgacgcaccc	6780
gaccgcgtt aacgcgtcct ggttgtcg ttagtgcgg tctgcacccc ggccgtcata	6840
gtcgctggca agcaccagca agtcgcggc gtcgtgcgg cgctgcagcc ccagcagacc	6900
actggtcagg tacttccagc gctgagggt gtacagcgcg ttgatggtgc cgttgtcgc	6960
gtcggcgttag ctcaaggccac gtggatccga cgtcttaccc ggcttctgca ccagcgggtc	7020
aaccaggcg tggtagcggt tgacccactg ggccgagtcg gtgcccagag ggcaggccgg	7080
cgagcggcg cagtccggcgg cgtagtcatgg gaaagcggtc tgaaatcccg ccattggct	7140
gatgtttcc tcgattgggc taacggctgg atcgatagcg cctgcgagga ccatcgcccg	7200
cacatgagta ccgaaccgtt ccaggttaagc ggtgcccac tcggtccgt agctgtatcc	7260
gaggtatgg atctgatcgt cacctaacgc ttggcgaacc atgtccatgt cccgtgcac	7320
ggacgcggta ccgatattgg ccaagaagct gaagccatc cggtaacac agtcctggc	7380
caactgccgg tagacctgtt cgacgtgggt gacaccggcc ggactgttagt cggccatcg	7440
atcgcccg taccgcgtcga actcggcgac ggtgcgacac cggtaacgcag ggtcgagtg	7500
gccgaccctt ctccggcgtga agcccaccag gtcgaagtgg cggagaatgt cggtgtcg	7560
gatcgccggt ccatacgccg cgaccatgtc gaccgcgcac gccccgggtc cccaggatt	7620
gaccagcgt gctccgaatc gctgtccgt cgcgggacg cggatcaccc ccaacttcgc	7680
tttgtccca ccgggttggt cgtagtcgac gggacggac accgtcgccg acgtcgact	7740
gcgaatttcg ctggtgtcgg cgtatgaactc cggcagactg ttccaaactct gttggcgc	7800
cacgaccggc gcacccgggg tttggccggc gccgggttct tcagtcgcgc cggcaacgg	7860
gggcgctgtc aggggcagtc cggcagacgca aacccgaag gacagcagcg cggagctcaa	7920
cggtgtcgcc cgccacatgg cggccatcgt ctcaccggcg aatacctgt acggcgcgaa	7980
atgatcacac cttcgttct tcgccccgt agcacttggc gccgctggc ggcgtgggtc	8040
cgccgattaa atacgcccgtc acgtactcgt caatgcagct gtcgcccgtgg aataccaccg	8100

tgtgctgggt	tccgtcgaag	gtcagcaacg	aaccgcgaag	ctggttcgcc	aggtcgaccc	8160						
cggccttgt	ta	cggcgtcg	cc	gggtcatgg	tggtg	gatac	caccaccgtc	ggcactaggc	8220			
cgggc	cccga	gacggcatgg	ggctgactt	g	gggtggcac	cggccaga	ac	gcgcagg	gtgc	8280		
ccagcggc	gc	atcaccgg	tg	aacttcc	cg	atc	gaa	cggtgc	gatc	8340		
ggcggt	t	tc	gtatgacc	ttgtc	cgat	cg	cgtaacc	gg	gctgat	cg	8400	
tcg	ccaccc	cg	cg	gaattgtt	gt	gcggcc	tg	cgag	tccc	gtac	8460	
tgtc	ggcc	ag	ccagcagg	gtgt	ctcc	gc	gat	tc	ac	gat	8520	
tca	agtgtt	cc	acagatt	c	gtg	ag	tata	at	gg	ccc	at	8580
tata	actc	ag	ccgc	cg	cc	tg	tc	gc	cc	gt	cg	8640
gg	tcgac	aa	cgatcg	acc	gg	ctgt	gg	tg	tt	gg	cc	8700
cc	agcgg	ca	gcccgtt	c	ttgg	cg	ca	gt	tt	gt	cc	8760
ag	ccctt	gg	ctgg	gc	tg	cc	tc	tg	cc	tc	tt	8820
cg	agaat	at	cat	tg	ccc	gc	acc	cg	at	cc	gg	8880
tg	ccgt	ac	ga	gt	ccc	agg	ta	gt	cc	gt	cc	8940
cc	aggt	tc	tt	gg	cc	tt	cc	ga	tt	tt	tt	9000
cc	acac	ag	cg	ac	tg	tt	gg	tc	cc	cc	gg	9060
tg	tagt	ca	ac	ct	gc	gg	tc	tt	cc	cc	gg	9120
tc	gccc	gg	cc	ccc	gc	gg	tc	tt	cc	cc	gg	9180
gc	accc	cg	ct	tt	gg	cc	ac	cc	tt	cc	gg	9240
gt	tc	cc	ac	cc	gg	cc	at	tt	cc	cc	gg	9300
gc	gc	cc	ac	gt	cc	gg	tc	tt	cc	cc	gg	9360
cg	cata	ac	gc	cc	gg	cc	at	tt	cc	cc	gg	9420
cc	acc	gg	ct	tc	gg	cc	ac	cc	cc	cc	gg	9480
ag	ccc	aca	ag	cc	gg	cc	ac	cc	cc	cc	gg	9540
tct	ttgt	cg	cg	cc	ca	at	gt	cc	cc	cc	gg	9600
gc	cc	ag	ac	cc	gg	cc	ac	cc	cc	cc	gg	9660
gc	ga	ct	cg	cc	gg	cc	ac	cc	cc	cc	gg	9720
tc	gacc	gg	at	cc	gg	cc	cg	cc	cc	cc	gg	9780
gc	gct	gg	cc	ac	gg	cc	cc	at	cc	cc	gg	9840
cg	at	cc	cc	at	cc	gg	cc	ac	cc	cc	gg	9900
tgt	ggcc	ca	at	cc	gg	cc	ac	cc	cc	cc	gg	9960
gca	agt	tg	ac	gt	cc	gg	cc	ac	cc	cc	gg	10020

accgcggccct cgccttgatc tggagtcaga acgcgtcacg cagccggta aaggcgtaac 10080  
ccatgctcga gcaaacatgc atgggctgag tggacgttc cagacacagc aactggcg 10140  
caggccactg agccgctgca tgcgcgatgg tatgccatg gggggcccg ggcgtctga 10200  
gggaaagaag tggcagactg tcagggtccg acgaacccgg ggacccta ac gggccacgag 10260  
gatcgaccccg accaccatta gggacagtga tgtctgagca gactatctat ggggccaata 10320  
ccccccggagg ctccggccg cggaaccaaga tccgcaccca ccacctacag agatggaagg 10380  
ccgacggcca caagtggcc atgctgacgg cctacgacta ttcgacggcc cgatcttcg 10440  
acgaggccgg catccggtg ctgctggtcg gtgattcgcc gccaacgctc gtgtacggct 10500  
acgacacccac cgtgccgatc tccatcgacg agctgatccc gctggtccgt ggcgtggtgc 10560  
gggggtcccc gcacgcactg gtcgtcgccg acctgcccgtt cggcagctac gaggccggc 10620  
ccaccggccgc gtggccgccc gccacccggt tcctcaagga cggcggccca catgcggta 10680  
agctcgaggg cggtgagcgg gtggccgagc aaatcgctg tctgacccgcg gcgggcattcc 10740  
cggtgatggc acacatcgcc ttcacccgc aaagcgtcaa caccttggc ggcttccgg 10800  
tgcaaggccg cggcgacgccc gccaacaaa ccatcgccga cgcgtcgcc gtcgccaag 10860  
ccggagcgtt tgccgtcgatg atggagatgg tgccgcga gttggccacc cagatcaccg 10920  
gcaagttac cattccgacg gtcggatcg ggcgtggcc caactgacgac ggccaggtcc 10980  
tggtatggca ggacatggcc gggttcagcg ggcacaagac cggccgttc gtcaaacggt 11040  
atgcccgtatgt cggtggtgaa ctacggcgatg ctgcaatgca atacgcccgg gaggtggccg 11100  
gcggggattt ccccgctgac gaacacagtt tctgaccaag cggatcagc cggatcgccg 11160  
ggcattgcgg tggcgccctg gatcgccgtc acgcccggatt gcccggccgg acgcccggc 11220  
gggaccatc ggcgtcgctg tcgcccgtt agccggggt gagcccagac attcgatgt 11280  
cccaacacca tccgcccacag cccaaattgtat gtggcactct atgcatgcct atcccccacc 11340  
aaccaccacc gcggcgacgc atcatgaccg gaggcgaaga tgccagtaga ggcgcccaga 11400  
ccagcgcgcc atctggaggt cgagcgcaag ttcgacgtga tcgagtcgac ggtgtcgccg 11460  
tcgttcgagg gcatcgccgc ggtgggtcg gtcgagcgt cggccacca gcagctcgac 11520  
gcgggtact tcgacacacc gtcgcacgac ctggcgccca accagatcac ttgcggccgc 11580  
cgccacccggcg gcggccgacgc cggctggcat ctgaagctgc cggccggacc cgacaacgc 11640  
accgagatgc gagcaccgct gtccgcattca ggcgacgctg tgccggccga gttgttggat 11700  
gtgggtctgg cgatcgcccg cgaccagccg gttcagccgg tcgcccggat cagcactcac 11760  
cgccaaagcc agatcctgta cggcgccggg ggcgacgcgc tggcggaaatt ctgcaacgac 11820  
gacgtcaccg catggtcggc cggggcattc cacggccgtg gtgcagcggaa caacggccct 11880  
gccgaacagc agtggcgcga atggaaactg gaactggta ccacggatgg gaccggcgat 11940

accaagctac tggaccggct agccaaccgg ctgctcgatg ccgggtgccgc acctgccggc 12000  
cacggctcca aactggcgcg ggtgctcggt ggcacctctc ccgggtgagct gccaacggc 12060  
ccgcagccgc cggcgatcc agtacaccgc gccgggtccg agcaagtcga gcagctgctg 12120  
ctgtggatc gggccgtgcg ggccgacgcc tatgacgccg tgcaccagat gcgagtgacg 12180  
acccgcaaga tccgcagctt gctgacggat tcccaggagt cgtttggcct gaaggaaagt 12240  
gcgtgggtca tcgatgaact gcgtgagctg gccgatgtcc tggcgtagc ccgggacgcc 12300  
gaggtactcg gtgaccgcta ccagcgcaa ctggacgcgc tggcgccgga gctggtacgc 12360  
ggccgggtgc gcgagcgccct ggtagacggg ggcggcgcc gataccagac cgggctgcgg 12420  
cgatcaactga tcgcattgcg gtcgcagcgg tacttccgtc tgctcgacgc tctagacgc 12480  
cttgtgtccg aacgcgccc tgcacttct ggggaggaat cggcaccggg aaccatcgat 12540  
gcggcctacc ggcgagtcgg caaagccgca aaagccgcaa agaccgcccgg cgaccaggcg 12600  
ggcgaccacc accgcgacga ggcattgcac ctgatccgca agcgcgcgaa gcgattacgc 12660  
tacaccgcgg cggctactgg ggcggacaat gtgtcacaag aagccaaggat catccagacg 12720  
ttgcttaggcg atcatcaaga cagcgtggtc agccggaaac atctgatcca gcagggcata 12780  
gccgcgaaca ccgcggcgaa ggacaccttc acctacggtc tgctctacca acaggaagcc 12840  
gacttggccg agcgtgccc ggagcagctt gaagccgcgc tgcgcaaact cgacaaggcg 12900  
gtccgcaaag cacgggattt ggcggccag gggcggacga gttggctgt aagccggatt 12960  
ctgttccgcg ccgcccacagc caagctaact gcggcacggc ggccgaccatc catctggaca 13020  
caccgttacc ggggtgcctcg agccgcctac ccgcaggctc gggcgagcaa ccctcaagcg 13080  
cctgcgcggc cgcactttcg gtgcggcctt cttggccttgc ctgcgggtgg gtttgccta 13140  
gccaccccg tcacccggaa tgctggcgctc ctcttaccgc accgtttcac cttgccacc 13200  
acgaggatgg cggctgttt tctgtggcac ttcccgcga gtcacctggg attgccgtt 13260  
gcaatcaccc tgctctgtga agtccggact ttccctcgact cgacgctgaa cctcgtaat 13320  
ccacacaaggc cctacgcgag ccgcggccgc ccagccaact catccgcgac gaccacgcta 13380  
ccccgctggg cgggtgcgcg gccagtgtga ccgcgtggacg acacggctag tcggacagcc 13440  
gatccggccg gcaagtcccta tcgtggactg gtgacacggg gggacaaacg cgtcgactcc 13500  
ggcgactggg acggcatcgcc tgccgaggctc agcgagtgacg gtggcgact gctacccgg 13560  
ctgatcaccc ccggcgaggc cgcccggtcg ccgaagctgt acgcccacga cggcctgttt 13620  
cgctcgacgg tcgatatggc atccaaggcg tacggcgccg ggcagtatcg atattccat 13680  
gccccctatc ccgagtgtac gaggctctca agcaggcgct gtatccaaa ctgctgccga 13740  
tagcgccaa ctggtgggcc aaactgggcc gggaggcgcc ctggccagac agccttgcgt 13800  
actgggtggc gagctgtcat gcccgggcc aaacccgatc cacagcgctg atgttgaagt 13860

acggcaccaa cgactggaac gccctacacc aggatctcta cggcgagttg gtgttccgc 13920  
 tgcagggtgt gatcaacctg agcgatccgg aaaccgacta caccggcggc gagttcctgc 13980  
 ttgtcaaca gcccgcctcgc gcccataatccc ggggtaccgc aatgcaactt ccgcaggac 14040  
 atggttatgt gttcacgacc cgtgatcggc cggtgccgac tagccgtggc tggtcggcat 14100  
 ctccagtgcg ccatgggctt tcgactattc gttccggcga acgctatgcc atggggctga 14160  
 tctttcacga cgcagcctga ttgcacgcca tctatagata gcctgtctga ttcaccaatc 14220  
 gcaccgacga tgccccatcg gcgtagaact cggcgatgct cagcgatgcc agatcaagat 14280  
 gcaaccgata taggacgccc gacccggcat ccaacgcccag ccgcaacaac attttgcgt 14340  
 gcgtgacatg tgacaccacc agcaccgtcg cgccttcgta gccaacgatg atccgatcac 14400  
 gtcccccccg aacccggcgc agcacgtcgt cgaagcttc cccacccggg ggcgtgatgc 14460  
 tggtgtcctg cagccagcga cggtgacgt cgggatcgcg ttctgccggcc tccgcaacg 14520  
 tcagccccctc ccaggcgccg aagtccgtcg cgcacccgtc gtcacgtcgt accacgtcca 14580  
 gggccagggc tctggcggcg gtcacccggc tgcgttaagc ccgtgttagc ggcgaggaga 14640  
 ccaccgcagc gatccccccg cgccgcgcca gataccggc ccgcgcacca acctggcgc 14700  
 acccccaccc gttcaacccc gggttgcgc gccccgataa gcggcggtgc tccgacagct 14760  
 ccgtctgccc gtggcgcaac aaaagttagtc ggggggtgt accgcggggc ccggtccagc 14820  
 cgggagatgt cggtgactcg gtcgcaacga ttttggcagg atccgatcc gccgcagccg 14880  
 attgcgcggc ggcgtccatc gcgtcatgg ccaacccgtc tgcatacgtg ttccgggcac 14940  
 gcggaaacca ctcgtagttg atccctgcgaa actggacgc caacgcctga gcctggacat 15000  
 agagcttcag cagatccggg tgcttgacct tccacccggc ggacatctgc tccaccacca 15060  
 gcttggagtc catcagcacc gcggccctcgg tggcacctag tttcacccggc tcgtccaaac 15120  
 cggctatcag gccgcggat tccggcgtat tgcgtcgacgt tgcgtcgcc ccggccgatc gcctgcttgg 15180  
 actcggccag cacgggtggag tgcgtcgcc tccacaccac cgcgcgtat ccggccggc 15240  
 cgggattgcc ccgcgcatccg ccgtccggctt cgtacacaac tttcactcc caaatccctc 15300  
 gagccgcac aagatcgctc cgcattccgg gcagcgcacc acttcatcc cggcggccgc 15360  
 cgagatctgg gccagctcgc cgcggccat ctcgtccgg caggcaccac atcgatgacc 15420  
 ttgcaaccgc ccggccctg gcccgcctcc ggccgcgtgt ctttcgtaga gccccgcaag 15480  
 ctcggatca agtgcgtcccg tcagcatgtc gcgtgcgtat gaatgttggt gcccggcttgc 15540  
 gtcgatttcg gcaagtgcct cgtccaaagc ctgcgtggcg gcggccaggt cggccgcac 15600  
 cgcttggagc gcccgcgact cggcggtctg ttgagcctgc agctccctcgc ggcgttccag 15660  
 cacctccagc agggcatctt ccaaactggc ttgacggcgt tgcaagctgt cgagctcgt 15720  
 ctgcagatca gccaatttgc tggcgccgt tgcacccgaa gtgagcaacg accggcccc 15780

gtcgccacgc ttacgcaccc catcgatctc cgactcaaaa cgcgacaccc ggccgtccaa 15840  
gtcctccgccc gcgattcgca gggccgccc cctgtcggt gcggcggtgt gctcgccctg 15900  
cacctgctgg taagccgccc gctgcccggcag atggtagcc cgatgcgcga tccgggtcag 15960  
ctcagcatcc agttcgcca attccagtag cgaccgttgc tgtgccactc cggcttcat 16020  
gcctgatctc tcccagtttc gtgatcgagg ttccacgggt cggtgcatat ggtgcacaca 16080  
cgcacccggca gcgacgcgccc gaaatgagac cgcaacactt cggcgccctg gccgcaccac 16140  
gggaattcgc ttgcccataatg cgcgacgtcg atcaggccca ttgcgaagc tcggcaatgc 16200  
tcgtcggtcg gatgatgtcg cagatcgccc gtaacgtacg cttgcacgtc cgcggcgccc 16260  
acggtgccaa gcaacgagtc cccggcgccc cccgacaccg cgaccgcga caccaggcagg 16320  
tcggatccc cggcgccgccc cacaccgtc gcagtcggcg gcaacgcggc ctccagacgg 16380  
gcaacaaagg tgcgccggcgg ttccgggttt ggcagtctgc caatccggcc taaccggctg 16440  
ccgaccggcg gtggtaccag cgcgaaatgc tcgaatccgg gtcctcgta agggtgcgcg 16500  
gcgcgcacatcg ccgccaacac ctcggcgccc gtcgtcggt gtgcgacgac ctcgacccgg 16560  
tcctcgccca cccgttcgac ggtaccgacg ctgcctatgg cgggcgacgc cccgtcgac 16620  
gccaggaact gcccggtacc cgcgacactc cagctgcagt gcgagtagtc gccgatatgg 16680  
ccggcaccgg cctcaaagac cgctgcccgc accgcctctg agttctcgcc cggcacatag 16740  
atgaccact tgcgagatc ggccgctccg ggcaccgggt cgagaacggc gtcgacggtc 16800  
agaccaacag cgtgtgcccac cgcgacgc acaccggcg acgcccggatc ggcgttggc 16860  
tgcgccgtaa acaacgagcg accggtccgg atcaggcggt gcaccagcac accctttggc 16920  
gtgttggccg cgaccgtatc gacccacgc agtaacaacg ggtggtgcac caatagcagt 16980  
ccggcctggg gaacctggtc caccaccgc ggcgtcgct ccaccgcaac ggtcaccgaa 17040  
tccaccacgt cgtcgggtc gccgcacacc agacccaccc aatcccacga ctggcaagc 17100  
cgccgggggt aggccctggtc cagcacgtcg atgacatcg ccagccgcac actcatcgcc 17160  
gtcctccacg ctttgcac tcggcgatcg ccgcaccagg cacggccac tccggcgca 17220  
ccgcccggcc caggtaccgc gcgtccaggc cgacgaagggt gtcaccggcgg cgacccgaa 17280  
ttccttgct ctgcaaatacg tttcgtaatc cgtcagcatc ggcgtatgtt aacagtagca 17340  
aaggggccgc accatcgacc acctcgac ccaccgatct cagtcggcc accatctccg 17400  
cgcgacgcgc cgtcaaccgc accgcacatcg ctgcggcagc ggcgaccggc cggggggcgc 17460  
agcaaggcgc gatggccgtc agttgcaatg ttcccaacgg ccagtgcgcgt cgctgcacgg 17520  
tcaaccgagc cagcacgtct ggccgagccga ggcgttagcc cacccgaat ccggccagcg 17580  
accacgtttt cgtcaagcta cggagcacca gcacatcgcc cagcgagtca tcggccaacg 17640  
attgcggctc gccgggaacc caatcagcga acgcctcgac gaccaccagg atgcgtcccg 17700

gccggcgtaa ctcgagcagc tgctcgccga ggtgcagcac cgaggtgggg ttggtcggat 17760  
 tacccacgac gacaaggctcg gcgtcgctcg gcacgtgcgc ggtgtccagc acgaacggcg 17820  
 gctttaggac aacatggtgc gccgtgattc cggcagcgct caaggctatg gccggctcgg 17880  
 tgaacgcggg cacgacgatt gctgcccga ccggacttag gttgtgcagc aatgcgaatc 17940  
 cctccgcccgc cccgacgagc gggagcactt cgtcacgggt tctgcatga cgttcagcga 18000  
 ccgcgtcttg cgcccggtgc acatcgctgg tgctcgata gcgccgcagc tccggcagca 18060  
 ggcggcggag ctgcccggacc aaccattccg gggccggtc atggcggacg ttgacggcga 18120  
 agtccagcac gcccggcgcg acatcctgat caccgtggta gcgcgcgcgc gcaagcggc 18180  
 tagtgtctag actcgccaca gcgtcaaaca gtagtggcc ggtgtgcggg ccaagaatcc 18240  
 agagcaccgc cgacgcgttg tctacgcggc gacaaccgcg acatcacagg cagctaacag 18300  
 ggcgtcggcg gtgatgatcg tcaggccaag cagctgtgcc tggcgtatga gcacacggc 18360  
 gaatggatgt cgatgggtat cccgaagctc tgccgtgcgc agtgtgtgcg tggtaactg 18420  
 acagcggcga cgtgcccgcag cggcgcattc gatcgggcac gtaagaagcc gatggctcg 18480  
 gccgcgggag cttgcccggg cggtagtgta tcgcgatctc ccaggcactg gcccgcaca 18540  
 agagaatgct gttgcggacg tcctgaacaa tcgcccgtgt ttcgttgacg gcatccgcag 18600  
 ccaaacgtgg gtgtcgatga ggtagcgcctt caccggtgaa agcgttcgag cacgtcgtct 18660  
 gacaacggag cgtccaaatc gtcgggcacg cggtagacgc catggtaat gcctaaccgc 18720  
 cgagtctcat gaggatgcag cggcacaaggc tttgctaccg gctgcgcgcg gcccgcata 18780  
 tcaacctctg cccgcgttag acgagccgca gcagctcgga caggcgtgtc ttgcgcgt 18840  
 gaacgcgcac cgcgttcgca ggcgcgcaga cttcgcgtc gaccacctgc tcaccaaact 18900  
 tcgcgatcat cgcctgatac cacagcgcac acggtagcg gttgtccaa cgcgttcgtc 18960  
 aacgacaatg ggatcgtgac cgacacgcacc gcgagcggga ccaattgccc gcctcccca 19020  
 cgcgcgcgcg cacggcgcgc acgtcgccg ggtgaatcgc cgcagctggt gatctcgat 19080  
 ctggacggca cgctgaccga ctgcgcgcgc ggaatcgat ccagcttccg acacgcgctc 19140  
 aaccacatcg gtgcggcactacagcgc cccgaaaggc gacctggca ctcacatcg cggccgcgc 19200  
 atgcatgaga cgctgcgcgc catgggcgc ggcgaatccg cggaggaggc gatcgtagcc 19260  
 taccggcccg actacagcgc cccggttgg gcgatgaaca gttgttcga cggatcg 19320  
 cccgtctgg cgcacctgcg caccggcggt gtccggctgg cgcgcgcac ctccaaaggca 19380  
 gagccgaccg cacggcgaat cctgcgcac ttcggattt agcagcactt cgaggtcatc 19440  
 gcgccgcgcg gcaccgatgg ctgcgcggc agcaaggctcg acgtgtggc ccacgcgctc 19500  
 gcgccagctgc ggccgcgtacc cgagcggttg gtgatggctcg gcgaccgcag ccacgcgctc 19560  
 gacggggcgg ccgcgcacgg catcgacacg gtgggttcg gctggggcta cggccgcgc 19620

gactttatcg acaagacac caccaccgtc gtgacgcatttccgacacat tgacgagctg 19680  
agggaggcgat taggtgtctg atccgctgca cgtcacattc gtttgtacgg gcaacatctg 19740  
ccggtcgcca atggccgaga agatgttgc ccaacagctt cggccaccgtg gcctgggtga 19800  
cgcgggtgcga gtgaccagtg cgggcaccgg gaactggcat gttaggcattt ggcggacga 19860  
gccccggcc ggggtgttc gagcccacgg ctaccctacc gaccaccggg cccgacaagt 19920  
cgccaccgaa cacctggcg cagacattt ggtggcattt gaccgcaacc acgctcgct 19980  
gttgcggcag ctcggcgtcg aagccgccc ggtacggatg ctgcggcat tcgaccacg 20040  
ctcgggaacc catgcgctcg atgtcgagga tccctactat ggcatcact ccgacttcga 20100  
ggaggtcttc gccgtcatcg aatccgcctt gcccggctg cacgactggg tcgacgaacg 20160  
tctcgcgcgg aacggaccga gttgatgccc cgcctagcgt tctgtctgcg gcccggctgg 20220  
ctggcgttgg ccctggcgtt ggtcgcttc acctacctgt gctttacggt gctcgcccg 20280  
tggcagctgg gcaagaatgc caaaacgtca cgagagaacc agcagatcag gtattccctc 20340  
gacaccccgcc cggttccgtt gaaaaccctt ctaccacagc aggattcgtc ggccggac 20400  
gcccggatggc gcccgggtgac ggcaaccgga cagtacattt cggacgtgca ggtgctggcc 20460  
cgactgcgcg tggggaggg ggaccaggcg tttgaggtgt tggcccccatt cgtggcgac 20520  
ggcggaccaa ccgtccttgtt cgaccgtgga tacgtgcggc cccaggtggg ctcgcacgt 20580  
ccaccgatcc cccgcctgcc ggtcagacg gtgaccatca cccggccgtt gctgtactcc 20640  
gaaccgagcg tggcgccaa agacccattt gtcagagacg gttccagca ggtgtattcg 20700  
atcaataccg gacaggtcgcc cgcgtcgacc ggagtccacg tggctgggtc ctatctgcag 20760  
ttgatcgaag accaaccggg cgggctcgcc gtgctcgccg ttccgcattt agatcccg 20820  
ccgttcctgt cctatggcat ccaatggatc tcgttcggca ttctggcacc gatcggttg 20880  
ggctatttcg cctacgcccga gatccggcg cggccgggg aaaaagcggg gtcgcacca 20940  
ccggacaaggc caatgacggt cgagcagaaa ctcgcgtgacc gtcacggccg cccgggtaa 21000  
accaacatca cggccaaatac cgcagcccccc gcctggacca cccgcgacag caccacggcg 21060  
cggcgcagat cggccacccctt gggcgaccgg cggcgtccca aggtggccg gatctgcac 21120  
tcatggtgtt accgggtggg cccacccacg cgcacgtcaa gtcacccagc aaacggcc 21180  
tcgacgacac cggcgttggg gctggatgg cggcgccgt cggccggccca gggccgtacc 21240  
gcaccgggg gcgacccacc gaccaccggc gtcacgtca ccaccagcac cggccgtcc 21300  
cgtgcgcacca catagttggc ccagtcatcc aatcggtctg cagcccaacc gaatggaga 21360  
taacgcggcg agcggtagcc gatcatcgag tccagggtgt tgatggcaccg atatccac 21420  
accgcaggca cggccgtcgat agccgcccac agcagcggca ccacccggc gtcggcggt 21480  
tttcggcca ccgactccag cggccgtcgat gtcaggcccg gggccggccag ctggggccgg 21540

tcacgccccgc acagcgacgg cagcagccgt cgccgcgcct cgacatcg tc gctccaaac 21600  
agg tccgata tctggcgcc ggtgcgcgcc agcgaagttc cgcccagcgc tgcccagg tg 21660  
gccgtcgccgg tggccgcccac gggccaggac ctgcccggta gccgctgcag tgccgcgcg 21720  
agcaaggccca ccgcgcgcac cagcaggccg acgtgtaccg caccggcgcac ccggccgtca 21780  
cggttaggtga tctgctccag cttggcgcc gcccggaccga acaggggccac cggatgac 21840  
cg tttggggt cgccgaacac gacgtcgagc agg cagccga tcagcacg cc gacggccctg 21900  
gtctgcagg tcgatgcaaa cactccggca gcgtcgac a cgtggtctac gctcagctat 21960  
ttatgacctc atacggcagc tatccacgat gaagcggcca gctacccggg ttgcccac 22020  
gttgaacccg gcccgaatgt ttttgcggc agcgaatgtc atcatgcagc tggcagtgc 22080  
gggtgtcggg tatggcgtgc tggaaagccc ggtggacagc ggcaacgtct acaagcatcc 22140  
gttcaagcgg gcccggacca ccggcaccta cctggcggtg gcgaccatcg ggacggaatc 22200  
cgaccgagcg ctgatccggg gtggcgtgg cgtcgccac cggcagggtc gtcgacggc 22260  
ctcgagccca gtgtcctata acgccttcga cccgaagttt cagctgtggg tggcggcgtg 22320  
tctgtaccgc tacttcgtgg accagcacga gtttctgtac ggcccactcg aagatgccac 22380  
cgccgacgccc gtctaccaag acgcca aacg gttagggacc acgctgcagg tgccggagg 22440  
gatgtggccg ccggaccggg tcgcgttcga cgagta ctttgg aagcgtcgc ttgtatgggct 22500  
gcagatcgac gcgcgggtgc gcgagcatct tcgcgggggtg gcctcggtag cgtttctccc 22560  
gtggccgttg cgccgggtgg ccggccgtt caacctgttt gcgacgacgg gattttggc 22620  
accggagttc cgccgcgtga tgcagctgg a tggtcacag gcccagcagc gtcgttcga 22680  
gtggttactt tccgtgtac gtttagccga ccggctgatt ccgcacatcg cctggatctt 22740  
cgtttaccag ctttacttgc gggacatcg gtttcgcgc c gacacggcc gccgaatcg 22800  
ctgatagagc ccggccgagt gtgagcctga cagcccgaca ccggcggcgt gtgtcgcc 22860  
gccaggttca cgctcgccga tctagagccg ccgaaaacct acttctgggt tgcctccga 22920  
atcaacgtgc t gatctgctc gagcagctca cgcataatcg cgcgcacatcg atccaccgc 22980  
gcatacaggt cggccttggc cgccggcagc tggccgcac tcattggccg caccggcgt 23040  
gctgtctgtc gcgcggcgtc gtgcgttga aacc caggtc gtcacccac gaccacgaca 23100  
ctgccatatac cggcgccccg ccgacaacga agcacagctt gcccggggc g cggacggga 23160  
tcgaaccggc gaccgctggt gtgtaaaacc agagctctac cgctgagcta cgcccccatt 23220  
accggccgag gtcacacgccc ttgcggccaa gcacccaaaa cttaggccg taagcgcgc 23280  
cagagcgtcg gtccacagcc gctgatcgac aacttcaccc ggctgcttca tctcgccgaa 23340  
ccgaatgatc cctgaccgat cgaccacaaa ggtgccccgg ttagcgtatgc cggcctgctc 23400  
gttgaagacg ccgtaggcct gactgaccgc gccgtgtggc cagaagtccg acaacagcgg 23460

aaacgtgaat ccgctctgct tcgcccagat cttgtgagtg ggtggcgggc ccaccgaaat 23520  
cgctagcgcg gcgcgtcggt cgttctcaaa ctggggcagg tgatcacgca actggtccag 23580  
ctcgccctgg cagatgcccgg tgaacgc当地 cggaaagaac accaacagca cgttcttgc 23640  
accccggttag ccgcgcagggg tgacaagctg ctgattctgg tcgcgc当地 tgaagtcagg 23700  
ggcggggctt ccgcacgttca gcatcagcgc ttgc当地 gccc gcgatccgg ctgtaccaat 23760  
ctgc当地 ggccgc tccagttgcc cagattgacc gacgaggctg gcatcagccc agctgtggc 23820  
gcgc当地 ctggc当地 caatctggc gggcaataca tggccggctt ggccggctt gggcgtcacc 23880  
acccaaatca caccgtcctc ggcgagcggg ccgc当地 ccatcagggt gtccaccaaa 23940  
tcgc当地 ctggc当地 catcacgcca ccacaacagg acgacatcga tgacctcgctc ggtgtcttca 24000  
tcgagcaact ctccccggca cgcttcttgc atggccgc当地 ggatgtcgctc gtccgggtct 24060  
tcgtccc当地 cccattcctg gataagttgg ttc当地 gggccatgtg 24120  
ttcgaggcgt gatccggc当地 gaccaccgtg gacccctt cagtc当地 gggccatgtg 24180  
cacaccgtcg cgatggc当地 tatcgtcgca cagccagaac cggccaccc gccc当地 cta 24240  
gaaggcggcc acgc当地 cattt tcaatgc当地 tgc当地 ttggc当地 gagcc gatcaacccg 24300  
ccgggtgaat tccgctgtcg acgc当地 cgc当地 accgatggca tttgccaccg cgc当地 cccgc当地 24360  
gtcgacatat gc当地 ttggc当地 gagccg catccccag ttgc当地 ggcc当地 acgc当地 cggcc当地 tc当地 gagactgccc 24420  
tgagaccgtc gaggcactgt tgc当地 ttggc当地 gagccg ggc当地 ttccgatgg tc当地 gggccggct 24480  
gttgc当地 gggcc当地 tgattgaacg cggccacgtt ggc当地 ttccgatgg cgtcccttgc当地 24540  
ggtggccggcc acgc当地 gtccac cgc当地 ggatgtcgcc aatcgc当地 ttggc当地 gagcc gatcaacccg 24600  
ctgc当地 gactcc cggatgtcg acgtcgccgc cgaagccgac accgacgc当地 acaccgacga 24660  
gc当地 ggtaggccc ggtgc当地 acgtcgccgc当地 cggccatccg atgagcacgg catgtgaggt 24720  
tccgacgatc cccatcagca gc当地 cgc当地 gagccgac gc当地 cccgc当地 ctc当地 ccttgg 24780  
gagctcccccc cc当地 ggc当地 ctgc当地 gagccacggc gc当地 cccatccg atgagcacgg catgtgaggt 24840  
tacctggctcg cagc当地 cggccaccg ggc当地 ttggc当地 gtccatccgca gaaccgagcg 24900  
gagtgccggct atccgccc当地 gacgc当地 gggtg cggccacgata gggggacgac catctaaaca 24960  
gc当地 acgcaaggc ggaagccgc当地 cacctacagg agtagtgc当地 tgaccaccga ttccgccc当地 25020  
cacgatctgg cccaaaactc aaacagc当地 agc当地 aaccgg accgagttcg ggtgatccgc当地 25080  
gagggtgtgg cgtcgatattt gccc当地 gagccgacatt gatcccgagg agaccccgatgg gtc当地 gggctggag 25140  
tcctttgaca cgc当地 tgc当地 gca acgctgc当地 ggcc当地 cccgctactt gatgttgc当地 25200  
ctgctagacg gggccggc当地 gca gggccggatgg gccc当地 cccgctactt gatgttgc当地 25260  
gtcaacacca tcccgaccgatgg gctggagccg tggttcccg ggc当地 acgaa gca gacgacgt 25320  
cgttatcgag cgtggatcgatgg atgaaatgc当地 gca gacgacgt 25380

ggtgtggcg tgggtggcca tatctcgacc tacgcgtcgt ccgcggcgct ctatgaggc 25440  
 ggtttcaacc acttcttccg cggcaagtgc cacccggcg gcggcgatca ggtgttcatc 25500  
 cagggccacg ctccccggg aatctacgcg cgcccttcc tcgaaggcg gttgaccgccc 25560  
 gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcgggtt gccgtcttat 25620  
 cccgaccccg gcgtcatgcc cgacttctgg gaattccca ccgtgtcgat gggtttggc 25680  
 ccgctcaacg ccatctacca ggcacggttc aaccactatc tgcatgaccg cggtatcaaa 25740  
 gacacctccg atcaacacgt gtggtgtttt ttggcgacg gcgagatgga cgaacccgag 25800  
 agccgtggc tggcccacgt cggcgcgctg gaaggcttgg acaacttgac cttcgatgatc 25860  
 aactgcaatc tgcaagcgact cgacggcccg gtgcgcggca acggcaagat catccaggag 25920  
 ctggagtcgt tcttccgcgg tgccggctgg aacgtcatca aggtggtgcg gggccgcgaa 25980  
 tgggatgccc tgctgcacgc cgaccgcgac ggtgcgcgtt gtaatttaat gaatacaaca 26040  
 cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccacttc 26100  
 ttcggccgcg acccagcgcac caaggcgctg gtggagaaca tgagcgacca ggatatctgg 26160  
 aacctcaaac gggcgccca cgattaccgc aagggttacg ccgcctaccc cgccggcgtc 26220  
 gaccacaagg gacagccgac ggtgatccgt gccaagacca tcaaaggcta cgcgctggc 26280  
 aagcatttcg aaggacgcaa tgccaccac cagatgaaaa aactgaccct ggaagacctt 26340  
 aaggagtttc gtgacacgcgac gcggtttccg gtcagcgacg cccagcttga agagaatccg 26400  
 tacctgccgc cctactacca ccccgccctc aacgccccgg agattcgatca catgctcgac 26460  
 cggcgccggg ccctcggggg ctttgttccc gagcgacca ccaagtccaa agcgctgacc 26520  
 ctgccgggtc gcgacatcta cgccgcgttgg aaaaaggctt ctgggacacca ggaggtggcc 26580  
 accaccatgg cgacgggtcg cacgttccaa gaagtgttgc ggcacaagca gatcgcccg 26640  
 cgatagtcgatcc cgatcattcc cgacgaggcc cgacacccgttggactc ctgggtttccg 26700  
 tcgctaaaga tctataaccg caatggccag ctgtataccg cggttgcacgc cgacctgtatc 26760  
 ctggcctaca aggagagcga agtcccccgatccatgc acggcatcaa cgaagccggg 26820  
 tcgggtggct cgatcgatccgc ggccggcacc tcgtatgcga cgccacaacga accgtatgc 26880  
 cccatttaca tcttctactc gatgttccgc ttccagcgac cccggcgatag cttctggcc 26940  
 gcggccgacc agatggctcg aggggttgcg ctcggggccca cccggcgccg caccaccctg 27000  
 accgggtgagg gcctgcaaca cgccgcacgttgc cactcgatccgc tgctggccgc caccaacccg 27060  
 gcgggtggctt cctacgaccc ggccttcgccc tacgaaatcg cctacatcgatggaaagcgga 27120  
 ctggccagga tggcgccggga gaaccccgag aacatcttct tctacatcac cgtctacaac 27180  
 gagccgtacg tgcagccgccc ggagccggag aacttcgatcccgaggccgt gctgcgggtt 27240  
 atctaccgct atcacgcggc caccgagcaa cgccaccaaca aggccgcagat cctggctcc 27300

ggggtagcga tgcccgccgc gctgcgggca gcacagatgc tggccgcga gtggatgtc 27360  
gccgcccacg tgtggtcggt gaccagttgg ggcgagctaa accgcgacgg ggtggccatc 27420  
gagaccgaga agctccgcca ccccgatcgg ccggcggcgc tgccctacgt gacgagagcg 27480  
ctggagaatg ctcggggccc ggtgatcgcg gtgtcggact ggatgcgcgc ggtccccgag 27540  
cagatccgac cgtgggtgcc gggcacatac ctcacgttgg gcaccgacgg gttcggttt 27600  
tccgacactc ggcccgccgc tcgcccgtac ttcaacacccg acgcccgaatc ccagggtggtc 27660  
gcggttttgg aggcggtggc gggcgacggc gagatcgacc catcggtgcc ggtcgccgc 27720  
gcccgcagt accggatcga cgacgtggcg gctgcgcggc agcagaccac ggatcccggt 27780  
cccggggcct aacgcccggcg agccgaccgc ctttggccga atcttccaga aatctggcgt 27840  
agcttttagg agtgaacgac aatcagttgg ctccagttgc cccgcggagg tcgcccgtcg 27900  
aactgctgga cactgtgccc gattcgctgc tgccgcgtt gaagcgtac tcggccggc 27960  
tggccaccga ggcagttcg gccatgcaag aacgggtgcc gttcttcgcc gacctagaag 28020  
cgccccagcg cgccagcgtg gcgcgtggc tgcaacgcgc cgtggtcaac ttctcgaaat 28080  
ggatgcacga cccgcacagt gacgtcggt ataccgcga ggcattcgag ctggtgcggc 28140  
aggatctgac ggcacggatc ggcgtgcgc agaccgtgga catggtcgg gtcaccatgg 28200  
agttcttcga agaagtcgtg cccctgtcg cccgttccga agagcgttg accgcctca 28260  
cggtggcat tttgaaatac agccgcgacc tggcattcac cgccgcacg gcctacgccc 28320  
atgcggccga ggcacgagggc acctgggaca gcccggatgga gcccacgcgtg gtggacgcgg 28380  
tggtaacgcgg cgacaccggc cccgagctgc tgtcccgccc ggccgcgcgt aattgggaca 28440  
ccaccgcgc ggcgaccgtg ctgggtggaa ctccggcgc cggtccaaat ggctccaaca 28500  
gcaacggcga cagcgagcgg gcaacggcagg atgtccgcga caccgcggct cgccacggcc 28560  
gcaacggcgt gaccgacgtg cacggcacct ggctggtggc gatcgctcc ggccagctgt 28620  
cgccaaaccga gaagttccctc aaagacactgc tggcagcatt cgccgacgc cccgtggta 28680  
tcggccccac ggcgccccatg ctgaccgcgg cgccacgc cgcgtacgc gcaatctccg 28740  
ggatgaacgc cgtcgcggc tggcgcggag cgccgcggcc cgtgtggct agggacttt 28800  
tgcccgaaacg cgccctgtatg ggcgacgcct cggcgtacgt ggccctgcgt accgacgtga 28860  
tgcggccccct agccgatgcc ggaccgacgc tcacgcgac gctagacgca tatctggatt 28920  
gtggcggcgc gattgaagct tgtgccagaa agttgttcgt tcacccaaac acagtgcgg 28980  
accggctcaa gggatcacc gacttcaccc ggcgcgtatcc caccgcacca cgccatgcct 29040  
atgtccctcg ggtggcggcc accgtgggtc aactcaacta tccgcacgcg cactgaagca 29100  
tcgacagcaa tgccgtgtca tagattccct cgccggcgtag agggggtcca gcagggggccc 29160  
cgaaaagata ccaggggcgc cgtcgacgg aaagtgtatcc agacaacagg tcgcggacg 29220

atctcaaaaa catagttac aggcccgtt tgggtttat atacaaaaac ctaagacgag 29280  
gttcataatc tgttacaccg cgcaaaaccg tttcacagt gttctttag acacgtgatt 29340  
gcgttgcctcg caccggaca gggtcgcaa accgaggaa tgggtcgcc gtggcttcag 29400  
ctgcccggcg cagcggacca gatcgccgatg tggtcgaaag ccgctgtatct agatcttgcc 29460  
cggttggca ccaccgcctc gaccgaggag atcaccgaca ccgcggtcgc ccagccattg 29520  
atcgctggcc cgactctgtt ggcccaccag gaactggcgcc gccgatgcgt gctcgccggc 29580  
aaggacgtca tcgtggccgg ccactccgtc ggcaaaatcg ccgcctacgc aatcgccgg 29640  
gtgatagccg ccgacgacgc cgatcgatgtt gcccacccc gcggcgccga gatggccaag 29700  
gcctgcgcca ccgagccgac ccgcattgtt ggggtgtcg gcggcgacga gaccgagggtg 29760  
ctgagtcgccc tcgagcagct cgacttggtc ccggcaaaacc gcaacgcccgc cggccagatc 29820  
gtcgctgccc gcccgtgac cgcgttggag aagctcgccc aagaccgcgc ggcaaggcg 29880  
cggttgcgtt cactgggtgt cgccggagcg ttccacaccc agttcatggc gcccgcactt 29940  
gacggcttttgc ccggccgcgc ggccaacatc gcaaccgcgc accccaccgc caccgtgttgc 30000  
tccaaccgcg acgggaagcc ggtgacatcc gcccgcgcgg cgatggacac cctggcttcc 30060  
cagctcaccc aaccggtgcg atggacactg tgcacccgcga cgctgcgcga acacacagtc 30120  
acggcgatcg tggagttccc ccccgccggc acgcttagcg gtatcgccaa acgcgaactt 30180  
cggggggttc cggcacgcgc cgtcaagtca cccgcagacc tggacgagct ggcaaaccta 30240  
taaccgcgga ctcggccaga acaaccacat acccgctagt tcgatttgcata cacaacat 30300  
tacgaaggga agcatgtgt gcctgtcaact caggaagaaa tcattgcgg tatcgccgag 30360  
atcatcgaaag aggttaaccgg tatcgagccg tccgagatca ccccgagaa gtcgttgcgtc 30420  
gacgacccgtt acatcgactc gctgtcgatg gtcgagatcg ccgtgcagac cgaggacaag 30480  
tacggcgatca agatccccga cgaggacccgc gcccgtctgc gtaccgtcg tgacgttgc 30540  
gcctacatcc agaagctcgaa ggaagaaaac ccggaggcgcc ctcaggcggtt ggcgcgcaag 30600  
attgagtcgg agaaccggc tgccgttgc aacgttcagg cgaggcttgc ggccgagtcc 30660  
aagttagtca gcctccacc gctaattggcg gtttccccag cgatgggttgc accggccgtca 30720  
cagcgacgac gtcgatctcg ccggacatcg agagcacgtg gaagggtctg ttggccggcg 30780  
agagcggcat ccacgcactc gaagacgagt tcgtcaccaaa gtgggatcta ggcgtcaaga 30840  
tcggcggtca cctcaaggat ccggatcgaca gcaacatggg ccgactcgac atgcgacgca 30900  
tgtcgatcg ccagcggatg ggcaagttgc tgggtcgaca gctatgggag tccggccggca 30960  
gcggggagggt cgatccagac ccgttgcgcg ttgttgcgg caccggatcta ggtggagccg 31020  
agaggattgt cgagagctac gacctgtatga atgcgggcgg ccccccggaaag gtgtcccccgc 31080  
tggccgttca gatgtatcg cccaaacgggtt ccgcggccgtt qatcggtctg cagcttgggg 31140

cccgccgcccgg ggtgatgacc ccgggtgtcgg cctgttcgtc gggctcgaa gcgatcgccc 31200  
acgcgtggcg tcagatcgtg atgggcgacg ccgacgtcgc cgtctgcggc ggtgtcgaag 31260  
gaccatcga ggcgctgccc atcgccgctgt tctccatgtat gccccatgcg tcgacccgca 31320  
acgacgagcc tgagcgggccc tcccgccgt tcgacaagga ccgcgacggc tttgtgtcg 31380  
gcgaggccgg tgcgctgatg ctcatcgaga cggaggagca cgccaaagcc cgtggcgcca 31440  
agccgttggc ccgattgctg ggtgccgta tcacccgttgc cgcctttcat atggggcgc 31500  
ccgcggccga tgggtttcggt gccggtaggg cgatgactcg ctgcgtggag ctggccgggt 31560  
tgtcgccggc ggacatcgac cacgtcaacg cgacggcac ggcgacgcct atcgccgacg 31620  
ccgcggaggc caacgcccattc cgcgtcgccg gttgtatca ggccgcgggt tacgcgccga 31680  
agtctgcgct gggccactcg atcgccgccc tcggtgcgt cgagtcgggt ctcacgggtc 31740  
tgacgctgctg cgacggcgtc atcccgccga ccctgaacta cgagacaccc gatcccgaga 31800  
tcgacccgttgc cgtcgccgc ggcgaaccgc gctatggcga ttaccgttac gcagtcaaca 31860  
actcggttcgg gttcggccggc cacaatgtgg cgcttgcctt cgggcgttac tgaagcacga 31920  
catcgccgggt cgcgaggccc gaggtggggg tccccccgt tgcggggcg agtccggaccg 31980  
atatgaaagg aacgttcgca agaccaatga cggagctgg taccggggaaa gcctttccct 32040  
acgttagtcgt caccggcatc gccatgacga ccgcgtcgcc gaccgacgcg gagactacgt 32100  
ggaagtgtt gctggaccgc caaagcgaaa tccgtacgt cgatgaccca ttcgtcgagg 32160  
agttcgacct gccagttcgcc atcgccggac atctgttgc ggaattcgac caccagctga 32220  
cgccggatcga actgcggccgg atgggatacc tgcagcgat gtccaccgtg ctgagccggc 32280  
gcctgtggga aaatgcccgc tcacccgagg tggacaccaa tcgattgtatg gtgtccatcg 32340  
gcacccgcct ggggtcggcc gaggaactgg tcttcgttgc cgcgtcgccg 32400  
gaatgaaggc ggtctcgccg ctgaccgtgc agaagtacat gcccaacggg gccggccgg 32460  
cggtcggtt ggaacggcac gccaaggccc ggggtatgac gccggatatcg gcgtcgcat 32520  
ccggcgccga gcccattcgcc cgtcggtggc agcagattgt gctggagag gccgtatcg 32580  
ccatctcgccg cgccgtggag accaggatcg aacgggtcc catcgccggg ttcgtcaga 32640  
tgcgcacatcgat gatgtccacc aacaacgacg accccgcgg tgcgtcgccg ccattcgaca 32700  
gggaccgcga cggctttgtg ttcggcgagg gcggccct tctgttgc gggaccgg 32760  
agcacgcac ggcacgtggc gccaacatcc tggccggat catgggcgcc agcatcacct 32820  
ccgatggctt ccacatggtg gccccggacc ccaacgggg acgcggccggg catgcgatta 32880  
cgccggcgat tcagctggcg ggcctcgccc ccggcgacat cgaccacgtc aatgcgcacg 32940  
ccacccggcac ccaggtcgcc gacctggccg aaggcaggac catcaacaac gccttggcg 33000  
gcaaccgacc ggcgggtgtac gcccccaagt ctgcctcgcc ccactcggtg ggcgggtcg 33060

gcgcggtcga atcgatcttg acgggtctcg cggtgcgcga tcaggtgatc ccggccgacac 33120  
tgaatctggtaaacacctcgat cccgagatcg atttggacgt ggtggcgggt gaaccgcgac 33180  
cgggcaatta cgggtatgcg atcaataact cgttcgatt cggcggccac aacgtggcaa 33240  
tcgccttcgg acggtaactaa accccagcgt tacgcgacag gagacctcgatgacaatca 33300  
tggcccccga ggcgggtggc gagtcgtcg acccccgcga tccgctgttgcggctgagca 33360  
acttcttcga cgacggcagc gtggaaatgc tgacgcgacg tgaccgcgtcc ggagtgcgtgg 33420  
ccgcggcggg caccgtcaac ggtgtgcga ccacgcgtt ctgcaccgac ggcaccgtga 33480  
tggcggcgcgcatggcgatc gaggggtgca cgcacatcgtaacgcctac gacactgcca 33540  
tcgaagacca gagtcccatac gtgggcatac ggcattcggg tggtgcccggtggctgaag 33600  
gtgtgcgggc gctgcacgcgtt gtaggcagg tggtcgaagc catgatccgc gcgtccggct 33660  
acatcccgca gatctcggtg tcgtcggtt tcgcccggg cggcgcgcacc tacggaccgg 33720  
cggtgaccga cgtcgtcgatc atggcgccgg aaagccgggt gttcgatcacc gggcccgacg 33780  
tggtgcgcag cgtcaccggc gaggacgtcg acatggcctc gtcgggtggccggagaccc 33840  
accacaagaa gtccggggtg tgccacatcg tcgcccgcga cgaactcgat gcctacgacc 33900  
gtggcgcccg gttggtcgga ttgttctgcc agcagggca tttcgatcgc agcaaggccg 33960  
aggccgggtga caccgacatc cacgcgtgc tgccggaaatc ctcgcgacgt gcctacgacg 34020  
tgcgtccgat cgtgacggcg atccctcgatc cgacacacacc gttcgacgag ttccaggcc 34080  
attggcgcc gtcgatggtg gtcgggtggc gtcgggtgtc gggtcgcacg gtgggtgtac 34140  
tggccaacaa cccgctacgc ctggcggtgcgcctgactc cgaaagcgca gagaaggcag 34200  
cgcggttcgt gggctgtgc gacgcgttcg ggattccgct ggtgggtggc gtcgatgtgc 34260  
cggttatct gcccgggtgc gaccaggagt ggggtggcgt ggtgcgcgt ggcgcacatc 34320  
tgctgcacgc gttcgccgag tgcaccgttc cgcgggtcac gtcggtcacc cgaaagaccc 34380  
acggcggggc atacattgcg atgaactccc ggtcggtgaa cgcgaccaag gtgttcgcct 34440  
ggccggacgc cgagggtcgcg gtgatggcgc ctaaggcgccgtcgacatc ctgcacaaga 34500  
agaagttggc cgccgctccg gagcacgaac gcgaagcgct gcacgaccag ttggcccgcc 34560  
agcatgagcg catcgccggc ggggtcgaca gtgcgttgaa catcggtgtg gtcgacgaga 34620  
agatcgaccc ggcgcataact cgacgcaagc tcaccgaggc gtcggcgcac gtcggccac 34680  
ggcgcggccg ccacaagaac atcccgctgt agttctgacc gcgacgacgc gcagaatcg 34740  
acgcgcgagg tccgcgcgt gcgattctgc gtctgcgtc cagttatccc cagcggtggc 34800  
tggtcaacgc gaggcgctcc tgcgtcgatc ggacgggtgcc taccgacgcgtcgtcc 34860  
tcgagaaggc cggcgggttc gccaccaccc cgcaattgtcgtcgtccgttgcgtcc 34920  
agctcgacgt ccaagtgaaa aacggcgcc tcgttcgttgcgttggatcgccgttgcgtcc 34980

cacaagagcc ggacctgttg ggccgcttgg cgctctcgat tggttcatg ggggggcacg 35040  
ccgtcgctg tctgggcacc gcccgcgt tgtatggatt cgacacggaa aacaccgtcg 35100  
ctatccatat gtcgatccc ggagtaagga tgcggccac ggtcggtctg atggtccacc 35160  
aaccgctgg tgcccggtc caacgggtgt caggtcgctc cgccgaccgcg cccgcatgga 35220  
ctgcccgttga ggtcgacgaa cagttgcgc gcccgcggc gctggccacc ctgcacgcgg 35280  
caactacggtc aatgcgctgc gtcgcagtg aaattgaaaa cgccgttgct gagcagcgag 35340  
gcccccggagg catcgctcgcg ggcgcgaaac tcttaccctt cgccgacgga cgccgcgaat 35400  
cgcccatgga gagcgagggt cggctcgta tgcgtcgacca cgggctgccc ttgcccgaac 35460  
ttcaataccca gatacacggc cacgggtggt aaatgtggcg agtcgacttc gcctggcccg 35520  
acatgcgtct cgccggccgaa tacgaaagca tcgagtggca cgcgggaccg gcggagatgc 35580  
tgcgcgacaa gacacgctgg gccaagctcc aagagctcggt gtggacgattt gtcccattt 35640  
tcgtcgacga tgcagacgca gacccggcc gcctggcgcc cccatcgcc cgccacctcg 35700  
accgcgcgca tatggccggc tgaccgctgg tgagcagacg cagagtcgca ctgcggccgg 35760  
cgcaatgcga ctctgcgtct gtcgcgctc aacggctgag gaactccta gccacggcga 35820  
ctacgcgctc gcatcccgt ggcaccagac cgatccgggt cggcggtcg agatatcg 35880  
ccacatccag cgccccctca tgggtcaccg cgtattcgaa ctccgccegg gtcacgtcg 35940  
tgcgtcgcc gaccggctcg gtggccgct cacatgtggc ggcggcagcg acgttggccg 36000  
cctcgcccccc gtaccgcgccc accagcgact cggcaatcc ggcggccgat ccggggccg 36060  
gcccagggtt cgccgggtcg ccgatcagcg gcagggtcg agtgcggcac ttgcggctc 36120  
gcagggtcg cagcgtgatg ggcgcattca gcacatcctc tgccatgttag cggattccg 36180  
tcagcttgcc gccgaccaca ctgatcacgc ccgacggcga ttcaaaaaca gcgtggtcac 36240  
gcgaaacgtc ggcgggtcg ccctggacac cagcaccggc ggtgtcgatt agcggccgca 36300  
atccccata ggcaccgatg acatccttgg tgccgaccgc cgtccccat ggcgggttca 36360  
ccgtatccag caggaacgtg atctctccg aagacgggttgg tggcacatcg ggaatcgcc 36420  
cgggtcgctc ttgcgtcgatc agcccgagat agatccggcc cagctgtcg ggcgtggcga 36480  
acacgaagcg gttcagctca ccggggatcg gaatggtcag cgcggcagtc ggattggcaa 36540  
acgacttcgc gtcgaagacc agatgtgtgc cgcggctggg gcgtagcctc agggacgggt 36600  
cgatctcacc cggccacacg cccgcccgt tgatgacggc acgcgcccac agcgcgaacg 36660  
actgcccgggt ggcgggtcg gtcaactcca ccgaagtggc ggtgacattt gacgcgcccc 36720  
cgtaagttag gatgcgggcg ccgtgtggg ccgcgggtcg cgcgacggcc atgaccagcc 36780  
gggcgtcgatc gatcaatttgc ccgtcgatcg cgagcagacc accgtcgagg ccgtcccccc 36840  
gaacgggtggg agcaatctcc accaccgtg acgcccggat tcggcgatcg cggggcaacg 36900

tcgcgcggc cgtacccgct agcacccgca aagcgtcgcc ggccaggaaa ccggcacgca 36960  
 ccaacgcccc cttggtgtga cccatcgacg gcaacaacgg gaccagttgc ggcatggcat 37020  
 gcacgagatg aggagcgttt cgtgtcatca ggattcccgcg ttcgacggcg ctgcgcgggg 37080  
 cgatccccac gttccgctg gccagatagc gcagaccgccc gtgcaccaac ttcgagctcc 37140  
 agcggctggt gccgaacgccc agatcatgct ttccaccaa ggccaccgtc agaccgcccc 37200  
 tggcagcatc taaggcaatg ccaacaccgg taatgccccc gcctatcacg atgacgtcga 37260  
 gtgcgccacc gtcggccagt gcggtcaggt cggcggagcg acgcgcgcgc ttgagtgcag 37320  
 ccgagtgccc catcagcaca aatatccgtt cagtgcgtgg gtaagttcgg tggccagcgc 37380  
 ggcggaatcg aggatcgaat cgacgatgtc cgcggactgg atggtcact gggcgatcag 37440  
 caacaccatg gtcggccagt gacgagcgtc gccggagcgc acactgccc accgctgcgc 37500  
 cactgtcagc cggcggccca acccctcgat caggacotgc tggctggtgc cgaggcgctc 37560  
 ggtatgtac accctggcca gctccgagtg catgaccgac atgatcagat cgtcaccgg 37620  
 caaccggtcg gccaccgcga caatctgctt taccaacgct tcccggtcgt ccccggtcag 37680  
 gggcacctcc cgcagcacgt cggcgatatg gctggtcagc atggacgcca tgatcgaccg 37740  
 ggtgtccggc cagcgacggt atacggtcgg gcccgcgcgccc gggcgatctc 37800  
 ggcaagtgtc accccggtcca cggcgtaatc gacgacgcag ctgcggctg ccccgaggat 37860  
 acgaccaccg gatccgcgc ggtcattact cattgacagc atgtgtataa ctgtaacgcg 37920  
 tgactcaccg cgaggaactc cttccaccga taaaatggga cgcgtgggg gatccgcgc 37980  
 cggccaagcc actttctgat ggcgtccggc cgttgctgaa gcagggtgtg ggcctagcgg 38040  
 actcgagca gcccgaactc gacccgcgc aggtgcagct gcgcggcgtcc gcccgtcgg 38100  
 gggcagacca 38110

<210> 25

<211> 2540

<212> DNA

<213> Homo sapiens

<400> 25  
 gaaaagggtgg acaagtccta tttcaagag aagatgactt ttaacagttt tgaaggatct 60  
 aaaacttggc tacatgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga 120  
 ttaaaaactt ttgctaattt tccaaatggc agtcctgttt cagcatcaac actggcacga 180  
 gcagggtttc ttataactgg tgaaggagat accgtgcggc gcttttagttt tcacgtcagct 240  
 gtagatagat ggcaatatgg agactcagca gttggaaagac acaggaaagt atccccaaat 300

tgcagattta tcaacggctt ttatcttcaa aatagtgccaa cgca	360
atccagaatgtcagtacaa agttgaaaac tatctggaa gcagagatca	420
gacaggccat ctgagacaca tgcagactat ctttgagaa ctggcaggt	480
tca	540
tcagacacca tatacccgag gaaccctgcc atgtattgtg aagaagctag	540
tttcagaact gcccagacta tgctcaccta accccaagag agttagcaag	600
tactacacag gtattggtga ccaagtgcag tgctttgtt gtgg	660
tggaaacctt gtgatcgtgc ctggcagaa cacaggc	720
tttttggcc ggaatcttaa tattcgaagt gaatctgatg ctgtgagttc	780
ttcccaaatt caacaaatct tccaagaaat ccatccatgg cagattatga	840
tttacttttggat atactcagtt aacaaggagc agcttgcaag agctggattt	900
tatgctttag gtgaaggtga taaagtaaag tgcttcact gtggaggagg	960
tggaaagccca gtgaagaccc ttggaaacaa catgctaaat ggtatccagg	1020
ctgttagaac agaagggaca agaataatata aacaatattc attaactca	1080
ttcacttgagtgtaaactac tgagaaaaca ccatcactaa ctagaagaat tgatgatacc	1140
atcttccaaa atcctatggt acaagaagct atacgaatgg gttcagttt caaggacatt	1200
aagaaaataa tggagggaaaa aattcagata tctggagca actataaattc acttgagtt	1260
ctggcag atcttagtga tgctcagaaa gacagtatgc aagatgagtc aagtca	1320
tcattacaga aagagattag tactgaagag cagctaaggc gcctgcaaga ggagaagctt	1380
tgcaaaatct gtatggatag aaatattgtc atcg	1440
acttgtaaac aatgtgctga agcagttagc aagtgtccca tgcgtacac agtcattact	1500
ttcaagcaaa aaattttat gtcttaatct aactctatag taggcattttt atgttgc	1560
tattaccctg attgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat	1620
tagcatttgc taccaagtag gaaaaaaaaat gtacatggca gtgttttagt tggcaatata	1680
atcttgaat ttcttgattt ttcaggat tagctgtatt atccattttt ttactgtta	1740
ttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtgt	1800
attcatagta tactgattta atttctaagt gtaagtgaat taatcatctg gat	1860
tctttcaga taggcttaac aaatggagct ttctgtatataatgtggag attagagtta	1920
atctcccaa tcacataatt tg	1980
gaaagataga gattgtttt agaggttgt tggtgttt taggattctg tccattttct	2040
tgtaaaggaa taaacacgga cgtgtgcgaa atatgtttgt aaagtgattt gcccattgttgc	2100
aaagcgtatt taatgataga atactatcga gccaacatgt actgacatgg aaagatgtca	2160
gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gtatgtctg agccagatca	2220

aagtatgtat gttgttaata tgcataagaac gagagatgg gaaagatata caccaaactg 2280  
 ttaaatgtgg tttctcttcg gggagggggg gattggggga ggggccccag aggggtttta 2340  
 gaggggcctt ttcactttcg actttttca ttttgttctg ttccggatttt ttataagtat 2400  
 gtagaccccg aagggtttta tgggaactaa catcagtaac ctaaccccg tgactatcct 2460  
 gtgctcttcc tagggagctg tgggtttcc caccaccac cttccctct gaacaaatgc 2520  
 ctgagtgctg gggcactttg 2540

<210> 26

<211> 103

<212> RNA

<213> Homo sapiens

<400> 26  
 agcuccuua acaaaagucu guugcuugug uuucacauuu uggauuuccu aauauaaugu 60  
 ucucuuuuua gaaaaggugg acaaguccua uuuucaagag aag 103

<210> 27

<211> 28

<212> RNA

<213> Homo sapiens

<400> 27  
 ggauuuccua auauaauguu cucuuuuu 28

<210> 28

<211> 1619

<212> DNA

<213> Homo sapiens

<400> 28  
 ccggccagatt tgaatcgccg gaccgttgg cagaggtggc ggcggcggca tgggtgcccc 60  
 gacgttgcgg cctgcctggc agccctttct caaggaccac cgcatctcta cattcaagaa 120  
 ctggcccttc ttggagggtc gcgcctgcac cccggagcgg atggccgagg ctggcttcatt 180  
 ccactgcccc actgagaacg agccagactt ggcccagtgt ttcttctgct tcaaggagct 240  
 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccgggtt 300

cgctttcctt tctgtcaaga agcagttga agaattaacc cttggtaat ttttggaaact 360  
 ggacagagaa agagccaaga acaaaaattgc aaaggaaacc aacaataaga agaaagaatt 420  
 tgagggaaact gccaagaaag tgcgcgtgc catcgacag ctggctgcca tggattgagg 480  
 cctctggccg gagctgcctg gtcccagagt ggctgcacca cttccagggt ttattccctg 540  
 gtgccaccag cttccctgtg ggccccttag caatgtctta ggaaaggaga tcaacatccc 600  
 caaatttagat gttcaactg tgctcctgtt ttgtcttcaa agtggcacca gaggtgcctc 660  
 tgccctgtca gcgggtgctg ctggtaacag tggctgcctc tctctctc tctcttttt 720  
 gggggctcat ttttgcgtt ttgattcccg ggcttaccag gtgagaagtg agggaggaag 780  
 aaggcagtgt ccctttgtt agagctgaca gcttgcgtcg cgtggcaga gccttccaca 840  
 gtgaatgtgt ctggacctca tggatgttgc gctgtcacag tcctgagtgt ggacttggca 900  
 ggtgcctgtt gaatctgagc tgcaggttcc ttatctgtca cacctgtgcc tcctcagagg 960  
 acagttttt tggatgttgc tttttttttt ggttagatgca tgacttgcgt 1020  
 gtgatgagag aatggagaca gagtccctgg ctccctact gtttaacaac atggcttct 1080  
 tattttgtttt gaattgttaa ttcacagaat agcacaact acaattaaaa ctaagcacaa 1140  
 agccattctta agtcattggg gaaacgggtt gaacttcagg tggatgagga gacagaatag 1200  
 agttagatgaa agcgtctggc agatactcct tttgccactg ctgtgtgatt agacaggccc 1260  
 agttagccgc gggcacatg ctggccgctc ctccctcaga aaaaggcgtt ggcctaaatc 1320  
 ctttttaaat gacttggctc gatgctgtgg gggactggct gggctgtgc agggctgtg 1380  
 tctgtcagcc caacccctcac atctgtcacg ttctccacac gggggagaga cgcagtcgc 1440  
 ccaggtcccc gctttctttt gaggcagcag ctcccgagg gctgaagtct ggcgttaagat 1500  
 gatggatttg attccctc cttccctgtca tagagctgca gggtggattt ttacagcttc 1560  
 gctggaaacc tctggaggc atctcggtc ttccctgagaa ataaaaagcc tgtcatttc 1619

<210> 29

<211> 27

<212> RNA

<213> Homo sapiens

<400> 29

ggcguacacac cuucggguga agucgcc

27

<210> 30

<211> 27

<212> RNA

<213> Homo sapiens

<400> 30

ggcgucacac cuucggguga agucgcc

27

<210> 31

<211> 12

<212> PRT

<213> Homo sapiens

<400> 31

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro  
1 5 10

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/11757

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C07H 21/02; G01N 27/26  
 US CL : 435/6; 536/23.1; 204/451

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
 U.S. : 435/6; 536/23.1; 204/451

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
 STN, EAST

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6329146 B1 (Crooke et al) 11 December 2001 (11.12.2001), column 40, example 11.	1
Y	US 5,807,682 A (Grossman et al) 15 September 1998 (15.09.1998), column 19, lines 2-18.	1
Y	US 6,355,428 (Schroth et al) 12 March 2002 (12.03.2002), column 8, lines 64-67.	1
Y	US 6,320,040 B1 (Cook et al) 20 November 2001 (20.11.2001), column 11, lines 14-22	1
Y	US 6,391,542 B1 (Anderson et al) 12 May 2002 (12.05.2002), column 36, example 18.	1

 Further documents are listed in the continuation of Box C. 

See patent family annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*B\* earlier application or patent published on or after the international filing date
- \*L\* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

\*E\* document member of the same patent family

Date of the actual completion of the international search

22 June 2002 (22.06.2002)

Date of mailing of the international search report

18 JUNE 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks  
 Box PCT  
 Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Jyothsna Venkat *J. Venkat*  
 11/17/02  
 11/17/02  
 JYOTHSNA VENKAT PH.D  
 Telephone No. (703) 308-1235

SI  
 11/17/02  
 EXAMINER  
 TECHNOLOGY CENTER 1600